Directed Evolution of G-Protein-Coupled Receptors for Expression and Stability

---

DISSERTATION

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der Universität Zürich

von

Igor Dodevski

aus

Laufen BL

Promotionskomitee

Prof. Dr. Andreas Plückthun (Vorsitz)

Prof. Dr. Amedeo Caflisch

Prof. Dr. Andreas Engel

Zürich 2010

Abstract

G-Protein coupled receptors (GPCRs) represent the largest superfamily of cell surface receptors in the human genome. They mediate the cellular responses to an enormous diversity of endogenous signaling molecules such as hormones and neurotransmitters, as well as environmental signals such as pheromones, smells, tastes and light. Because of the enormous diversity in transmitted signals, GPCR signaling is involved in nearly every physiological process, and deregulated signaling readily leads to pathologic conditions. Currently, about 25% of all drugs in clinical use exert their action by targeting GPCRs.

Understanding the molecular mechanisms of ligand binding and signal transduction at the structural level of single atoms holds great promise for the detailed characterization of GPCR function and structure-based drug discovery. However, functional characterization and effective drug design of these receptors is strongly restricted by the limited availability of high-resolution structural information. This limitation is mainly a consequence of the enormous experimental difficulties inherent to the process of atomic-resolution determination of GPCR structure.

This thesis describes the development and application of powerful directed evolution methods aimed at improving the biophysical properties of GPCRs to make these receptors amenable to atomic-resolution structural studies. The initial studies have focused on engineering a water-soluble analog of the β_2 adrenergic receptor (ADRB2) by a complete redesign of its hydrophobic phospholipid-contacing surface. Computational methods were used to design a combinatorial gene library from which water-soluble ADRB2 analogs can be selected by modern in-vitro selection technologies and further evolved towards the desired biophysical and functional properties.

The more recent studies presented in this thesis describe the development of a powerful directed evolution method for producing well-expressed and stable GPCRs in the inner membrane of the expression host *Escherichia coli*. The method is based on introducing random amino acid changes to a GPCR sequence and selecting improved receptor variants by fluorescence-activated cell sorting. The performance of the method was first critically assessed on the model protein neurotensin receptor NTR1 and then successfully applied to three additional human GPCRs – some of which were hardly expressed or quickly lost their native fold when solubilized in detergent micelles. For all four receptors, mutations could be evolved that showed strongly increased functional expression levels (up to 30-fold) and greater receptor stability. The

improvements are achieved by a combinatorial approach that does not require the input of rational design principles.

In summary, by introducing the concepts of directed evolution to the field of integral membrane proteins, this work has produced powerful protein design strategies for GPCRs and a set of well-behaved GPCR proteins that are amenable to crystallographic studies. Existing roadblocks in structural and biophysical studies of these important receptors can now be removed by providing sufficient quantities of correctly folded and stable receptor protein. The ability to obtain atomic-resolution structures may help to elucidate the molecular basis for activation, inactivation, or pathology associated with that receptor, and may also provide valuable templates for drug design.

Zusammenfassung

G-protein gekoppelte Rezeptoren (GPCRs) stellen die grösste Protein-Superfamilie im menschlichen Genom dar. Deren Aufgabe ist es, äussere Signale, z. B. von Hormonen oder Neurotransmittern aber auch Sinneswahrnehmungen wie Geruch oder Licht, ins Zellinnere weiterzuleiten und in der Zelle eine physiologische Antwort auszulösen. Aufgrund ihrer Fähigkeit eine Vielzahl äusserst unterschiedlicher Signale zu verarbeiten, hängen die meisten physiologischen Prozesse im menschlichen Körper von der Aktivität der GPCRs ab. Gerät eine bestimmte GPCR-abhängige Signalverarbeitung jedoch ausser Kontrolle, können leicht pathologische Prozesse ausgelöst werden. Dies macht GPCRs zu prominenten Kandidaten in der pharmazeutischen Industrie und erklärt auch weshalb jedes vierte Medikament auf dem Markt in die Aktivität GPCR-abhängiger Signalverarbeitung eingreift.

Wäre man nun in der Lage, bestimmte funktionelle Aspekte G-protein gekoppelter Rezeptoren auf der strukturellen Ebene einzelner Atome genauer beschreiben zu können (z.B. die räumliche Konfiguration einer Bindesstelle für Liganden), ergäben sich daraus nicht nur wertvolle Erkenntnisse für die Grundlagenforschung sondern auch neue Möglichkeiten für die strukturgestützte Medikamentenentwicklung. Detaillierte strukturelle Einblicke sind heute jedoch nur sehr beschränkt möglich, weil trotz grosser Anstrengungen nur einige wenige hochaufgelöste Kristallstrukturen von GPCRs gemessen werden konnten. Diese Beschränkung besteht vor allem auf Grund der grossen experimentellen Schwierigkeiten bei der Röntgenstrukuranalyse von GPCRs.

Die vorliegende Arbeit beschreibt die Entwicklung und Anwendung neuer Methoden auf dem Gebiet des *Protein Engineering*, die die biophysikalischen Eigenschaften von GPCRs verbessern und damit die Strukturanalyse dieser Proteine ermöglichen sollen. Der erste Teil dieser Doktorarbeit hatte zum Ziel wasserlösliche Analoga des β_2-adrenergen Rezeptors zu konstruieren, indem der problematische wasserabweisende Anteil der Proteinoberfläche stark verringert wurde. Mittels computergestüzter Analysen wurde eine kombinatorische Genbibliothek des Rezeptors entworfen. Aus dieser sollten mit Hilfe von in-vitro Selektionsmethoden funktionale wasserlösliche Analoga des β_2-adrenergen Rezeptors selektiert werden können.

Der zentrale Teil dieser Doktorarbiet beinhaltet die neueren Studien zur Entwicklung einer effizienten Selektionsmethode zur Herstellung hoch exprimierender und stabiler GPCRs im Expressionsorganismus

Escherichia coli. In dieser Methode wird die Aminosäuresequenz eines experimentell schwierig zu handhabenden GPCR zuerst nach dem Zufallsprinzip mutiert. Mit Hilfe eines Durchflusszytometers können dann diejenigen Mutanten isoliert werden, die eine erhöhte funktionale Proteinexpression aufweisen. Die Effizienz der Methode wurde zuerst am Modellprotein Neurotensinrezeptor NTR1 gezeigt und danach erfolgreich auf drei weitere humane GPCRs angewendet. Für alle vier Ausgangsmoleküle konnten neue Varianten evolviert werden, die eine stark verbesserte funktionale Proteinexpression (bis zu 30 mal höher als das Ausgangsmolekül) und eine höhere Proteinstabilität aufwiesen.

Zusammenfassend resultierten aus dieser Arbeit neue effiziente Strategien für das *Protein Engineering* von G-protein gekoppelten Rezeptoren. Weil die evolvierten Proteine sehr stabil sind und in ausreichender Menge hergestellt werden können, entfallen nun einige der grössten Hürden im Prozess der Röntgenstrukturanalyse von GPCRs. Die Entdeckung neuer Kristallstrukturen von GPCRs würde dadurch vereinfacht und könnte dazu beitragen, die molekularen Mechanismen der Rezerptor-Aktivierung, der Rezeptor-Inaktivierung oder pathologischer Zustände besser zu verstehen. Auch die Entwicklung verbesserter Medikamente mit Hilfe strukturbasierender Methoden würde durch detaillierte Rezeptorstrukturen vorangetrieben.

Table of Contents

Chapter 1

Introduction and scope of thesis

1.1 Structure and function of G-protein coupled receptors

G-Protein coupled receptors (GPCRs) are cell surface receptors that mediate the cellular responses to an enormous diversity of endogenous signaling molecules such as hormones and neurotransmitters, as well as environmental signals such as pheromones, smells, tastes and light (**Fig. 1.1**). The list of ligands detected by members of the GPCR family range from peptides and protein hormones (e.g. angiotensin, bradykinin, neurotensin, substance P), biogenic amines (e.g. adrenaline, dopamine, histamine, serotonin) and lipids and eicosanoids (e.g. cannabinoids, leukotrienes, prostaglandines, thromboxanes), nucleotides and nucleosides (e.g. adenosine, ADP, ATP, UTP) and a wide variety of other molecules down to Ca^{2+} ions and photons (light). Binding constants are available for several thousand ligand-receptor combinations[1,2].

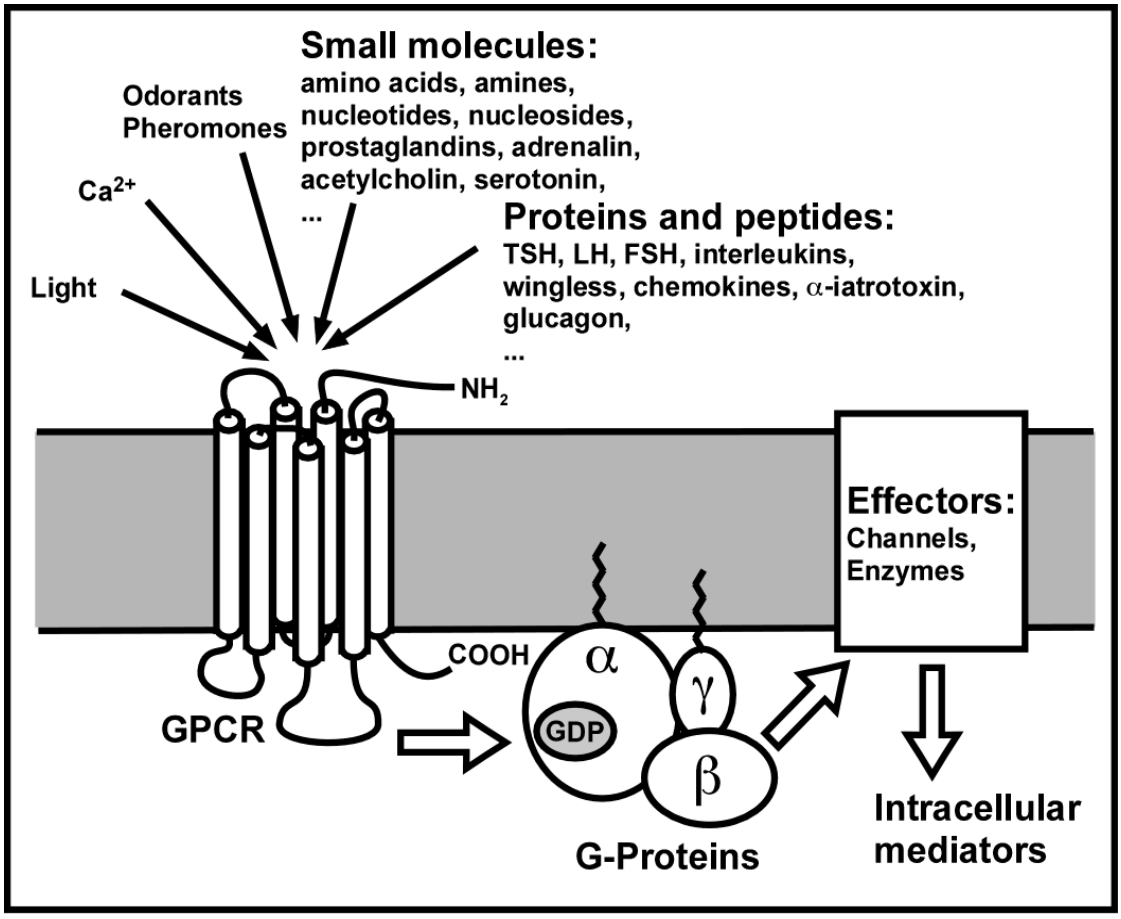


Figure 1.1: Schematic representation of the signaling mechanism of G-protein coupled receptors.

Despite the chemical and functional diversity of the signaling molecules recognized by the GPCRs, all of the G-protein coupled receptors are related on the sequence level, have the same membrane topology and are predicted to have a similar three-dimensional structure. They consist of a single polypeptide chain that threads back and forth across the phospholipid bilayer of the cell membrane seven times, with the N-terminus located on the exterior side of the plasma membrane and the C-terminus on the cytoplasmic side. From this conserved membrane topology stems their alternative name of "seven transmembrane domain receptors" (7TM).

The G-protein coupled receptors constitute one of the largest protein superfamilies. The *Caenorhabditis elegans* genome encodes ~1100 GPCRs (5% of all open reading frames), while in vertebrates the GPCR superfamily contains 1000-2000 members (>1% of the open reading frames)[3]. In the human genome approximately 800 GPCRs have been found, of which half are olfactory receptors, while half are thought to represent receptors for endogenous ligands[3,4]. The latter group can be subdivided into five sequence families according to the GRAFS classification system, which is strictly based on phylogenetic criteria[5]. The largest of theses families – the rhodopsin family – contains about 270 different receptors. An alternative name commonly seen for this family is Class A GPCRs[6,7]. Despite the very low sequence identity found among the members of the rhodopsin family, there are several sequence characteristics that indicate the common evolutionary decent of this family. These characteristics include the D(E)-R-Y(F) motif in TM3, the NSxxNPxxY motif in TM7, and several additional highly conserved amino acids found in each transmembrane helix of the 7TM fold.

All GPCRs share a common signaling mechanism, utilizing heterotrimeric G-proteins[8] (guanine nucleotide binding proteins) as second messengers (**Fig. 1.1**). The signaling mechanism of these heterotrimeric G-proteins depends on the dissociation of their Gα subunit from the Gβγ subunit complex, stimulated by the exchange of guanine nucleotide diphosphate for guanine nucleotide triphosphate on the Gα subunit. The complex of the Gβγ subunit can then modulate the activity of various effector proteins such as membrane channels and enzymes, e.g. adenylyl cyclases, phosphodiesterases, phospholipases, Ca^{2+} and K^{+}-channels.

Over the last years it has become apparent that GPCR mediated signaling goes beyond the classical mechanisms involving G-proteins. Numerous proteins have been found to interact with GPCRs, which may help to explain the remarkable diversity of cellular responses mediated by

these receptors. The effects of these proteins may range from small refinements of a transmitted signal to large shifts in signal trafficking. A prominent example are the β-arrestins. They have been found to be involved in several regulatory processes, such as receptor desensitizing and internalization. But they also serve as scaffolding proteins to facilitate the formation of G-protein independent signaling complexes[9]. There are many other examples of GPCR-activated signaling cascades that do not depend on G-protein recruitment[10].

Because of the wide variety of transmitted signals, GPCRs are involved in nearly every physiological process and deregulated signaling readily leads to pathologic conditions. This makes these receptors prime targets for drug development. Presently, about 25% of all known drugs in clinical use target GPCRs[11]. But these drugs hit merely 10% (corresponding to about 30 different receptors) of the pharmacologically relevant (non-olfactory) GPCRs. The remaining majority of non-olfactory GPCRs are considered potential high-value drug targets, and exploring their therapeutic potential remains a major challenge in academic and industrial GPCR research.

Understanding the molecular mechanisms of ligand binding and signal transduction at the structural level of single atoms holds great promise for the detailed characterization of GPCR function and protein-based drug discovery. The molecular structure of GPCRs specifies their chemical interaction with diffusible ligands and intracellular effector proteins and ultimately defines their role in physiology and pathology. However, functional characterization and effective drug design of these receptors is strongly restricted by the limited availability of high-resolution structural information. This limitation is mainly a consequence of the enormous experimental difficulties inherent to the process of atomic-resolution determination of GPCR structure.

1.2 Experimental difficulties in biochemistry and structural biology of GPCRs

The structural and functional study of integral membrane proteins is considerably hampered by the hydrophobic nature of these proteins. Problems commonly appreciated include low expression levels, toxicity to cells forced to overexpress these molecules, limited stability and aggregation in detergent-solubilized form, and the inability to refold them from solubilized non-native aggregates in vitro. A protein can only be crystallized and its atomic-resolution structure solved if it can be produced in sufficient amounts in soluble form and if it retains its native structure upon solubilization.

Because of the experimental difficulties, only about 210 unique structures of membrane proteins have been solved so far[12], as compared to ten thousands of structures of soluble globular proteins. Many of these represent proteins that are highly expressed in their natural source tissues or organelles, such as the visual cone pigment rhodopsin, the photosynthetic reaction center, photosystems and components of the respiratory chain. Other membrane proteins show a high intrinsic biophysical stability, which favors their functional solubilization in detergent micelles. Examples include β-barrel porins and α-helical ion channels. Yet others, such as the monotopic membrane proteins or α-helical transporters, contain relatively large intracellular and extracellular domains, minimizing the fraction of hydrophobic surface that needs to be masked with detergents to solubilize the protein. However, most membrane proteins, such as the members of the GPCR superfamily, are very difficult to express, rapidly unfold in detergent micelles and lack large hydrophilic surfaces for forming crystal contacts.

While GPCRs can be recombinantly expressed in eukaryotic cells at the low expression levels needed for binding studies and for the investigation of receptor signaling functions, so far most attempts to produce GPCRs in soluble form in amounts and quality suitable for crystallization have failed. A breakthrough in high-resolution X-ray crystallography was the structure determination of the bovine visual cone pigment rhodopsin at 2.8 Å in the year 2000[13]. These structural studies were made possible because the protein can be functionally purified in large amounts from bovine retinal tissue and because it can be maintained in its relatively stable non-signaling conformation. In recent years four more GPCR structures could be solved: the squid rhodopsin[14], the human β_2 adrenergic receptor[15,16], the turkey β_1 adrenergic receptor[17], and the human adenosine receptor A_{2A}[18]. These proteins, with the exception of squid rhodopsin, had to be overexpressed in a heterologous expression

system, the baculovirus system. A handful of additional GPCRs can be recombinantly expressed and purified in functional form at levels necessary for structural studies. These include the human histamine receptor H_1 expressed in the baculovirus system[19], and the rat neurotensin receptor NTR1 and the human cannabinoid receptor CB_2 both expressed in *Escherichia coli*[20,21]. For most GPCRs, however, finding satisfactory overexpression conditions remains very challenging.

Establishing a strong overexpression system represents only the first experimental obstacle towards the production of large amounts of a GPCR protein suitable for structural studies. Unlike soluble globular proteins, integral membrane proteins absolutely require the help of amphipathic detergent molecules to extract them from their natural physical environment – the phospholipid membrane – and to transfer them into a soluble state established by detergent micelles. This process is problematic from a thermodynamic point of view because the physical environment provided by the detergent micelle is very different from the phospholipid membrane. Protein solubilization by detergents therefore frequently leads to protein unfolding, aggregation, and loss of function.

For GPCRs the solubilization process turns out to be particularly destructive because of their intrinsic biophysical instability. GPCRs naturally exist as structurally flexible molecules to exert their function in transducing extracellular signals across the phospholipid bilayer[22]. The observation that GPCRs can activate G-proteins even in the absence of an activating agonist ligand, a phenomenon called basal activity, implies that the receptors can continuously adapt different thermodynamic receptor conformations. This structural and thermodynamic heterogeneity is most likely responsible for the rapid receptor unfolding observed upon receptor solubilization with detergents.

Crystallographic studies remain very challenging even if a protein can be overexpressed in large amounts and solubilized in its active form. For integral membrane proteins, crystal contacts are normally formed by amino acids of the polar surface area of the protein and not by the hydrophobic surface, which is masked by detergent molecules. In many cases however, the polar surface area represents only a minor fraction of the total surface area of the protein-detergent particle. Therefore, the likelihood of forming stable crystal contacts is strongly limited and obtaining diffraction-quality crystals remains very difficult. There are some tricks used to overcome the difficulty offered by the crystallization of proteins that have much of their molecular surface masked by detergents. Among them is the increase of the hydrophilic surface by co-crystallizing the membrane protein as a complex with an antibody fragment[15,23-25] or with a designed ankyrin repeat protein[26] or by

expressing the protein as a fusion to a highly soluble and well-folding domain[16,18].

1.3 Current strategies for overcoming the difficulties in the production of integral membrane proteins

For improving the process of producing a membrane protein at sufficient amounts and quality for crystallization, an ever-growing number of biochemical techniques are being developed and implemented. The high pace of the methodological development most probably reflects the observation that experimental parameters need to be optimized individually for any given target protein.

Currently there are three different strategies for increasing the chances of producing a well-expressed, soluble, and active membrane protein. The first strategy, which is the most intuitive and most widely used one, relies on the extensive screening of a large number of experimental conditions for expression and solubilization of a given construct. Typically, these conditions have to be optimized for any given target protein. A second strategy relies on the screening of protein homologs in order to identify a candidate that shows biophysically favorable properties. The homologs are usually screened in the context of a few globally optimized conditions for expression and solubilization. A third strategy involves the techniques of protein engineering to improve the biophysical properties of a target protein itself that otherwise fails to be overexpressed and solubilized. While the first strategy aims at the optimization of extrinsic experimental parameters, the latter two strategies focus on the intrinsic biophysical properties of the target protein itself. In the following the three strategies are discussed in more detail.

1.3.1 The classical strategy: Screening of extrinsic experimental parameters

The most common approach for finding suitable overexpression conditions for a given membrane protein is the systematic examination of a variety of different expression parameters. These parameters typically include the screening of expression hosts (prokaryotes and eukaryotes), promoters, N-terminal and C-terminal fusion adducts, expression media, and expression temperature. For the functional expression of GPCRs in the versatile host *E. coli* a combination of the following expression parameters has been found beneficial for some receptors: A genetic fusion of maltose binding protein to the receptor N-terminus and a thioredoxin A fusion to its C-terminus, combined with a low expression temperature (e.g. 20 °C). However, despite the fact that large resources

are invested in the process of optimizing expression parameters, many membrane proteins, especially those of eukaryotic origin, fail to be overexpressed in functional form.

In a recent study on the expression of GPCRs in bacteria it has been shown that functional expression can be improved by genetic engineering of the expression host. Co-expression of additional copies of the membrane-bound AAA+ protease FtsH, enabled the efficient production of the human cannabinoid receptor CB_1 in *E. coli*[27]. A similar effect was obtained by inactivating the chaperone DnaJ by a Tn5 insertion[28]. Genetically engineered expression hosts may therefore enable the production of membrane-integrated human GPCRs that are otherwise difficult to express, but the general applicability of this strategy remains to be demonstrated.

Cell-free production of functional GPCRs by the help of so-called nanodisks has been shown to be possible for several receptors[29-31]. If this technology proves to be generally applicable to integral membrane proteins, it may represent a very interesting alternative to cell-based expression systems.

If sufficient amounts of a membrane protein can be expressed, the next step that has to be optimized before initiating crystallization trials is to purify the protein. For the purification process the protein has to be handled in solution. Therefore, solubilization conditions have to be established that allow the protein to be efficiently extracted from the phospholipid bilayer and maintained in a functional form in solution. The choice of a detergent that satisfies such conditions cannot be made rationally, because the molecular interactions between protein and detergent are not well understood. Rather, finding an appropriate detergent for any given protein relies on the screening of many different detergents[32]. As for the screening of expression conditions, detergent screens are very laborious and many proteins fail to be solubilized in functional form.

Interestingly, if the protein is intended for crystallization, the choice of the detergent seems to be strongly limited. A systematic analysis of the crystallization conditions of solved structures of α-helical membrane proteins shows that only a small group of detergents conform to the formation of diffraction quality crystals[33]. In 90% of the successful crystallizations the detergents were alkyl maltopyranosides, alkyl glucopyranosides, or polyoxyethylene glycols. This implies that if a membrane protein fails to be functionally solubilized in one of these detergents, the protein will be very difficult to crystallize.

1.3.2 The genomic strategy: Screening of homologs

If the extensive screening of extrinsic experimental parameters fails, a second strategy referred to as *homolog screen* can be employed for increasing the chances of obtaining a well expressed and properly behaved membrane protein. With the advent of the genomics era the data of hundreds of fully sequenced genomes have become available, which allows the rapid cloning of many homologous sequences of a protein of interest. A set of homologs can then be screened in parallel to identify the most promising candidates for crystallization trials. When homologs are screened, other experimental parameters are usually kept rather constant. Therefore, the homolog screen usually involves a few globally optimized expression conditions and a small set of detergents, which are likely to promote crystal contact formation. The crystal structures of many membrane proteins, mostly of prokaryotic origin, could be solved by this strategy. However, if the research goal is to solve the structure of a particular membrane protein – for example a particular human GPCR – screening of homologs is not a suitable option. For proteins such as GPCRs, there are no bacterial homologs.

In order to accelerate screening of large sets of homologs (or a large variety of expression conditions for that matter), new technologies are continuously being developed. For example, fusing GFP to a membrane protein has been shown to be a good indicator for productive overexpression in *E. coli* and this system has been successfully implemented in many screens[34,35]. GFP fluorescence can also be used for quality control of membrane protein folding[36].

1.3.3 The protein engineering strategy: Improving the biophysical properties of the protein

The third strategy for improving the production of a membrane protein is based on the assumption that the biophysical properties of the protein itself can be improved by modifying the underlying amino acid sequence. In the research field of soluble globular proteins, biophysical properties, such as protein expression, stability, or solubility, are routinely improved by a strategy commonly referred to as protein engineering. However, most of the highly efficient protein engineering methods for soluble proteins cannot be applied directly to integral membrane proteins, due to the hydrophobic nature of these proteins. Developing protein engineering methods for integral membrane proteins represents the basis of the present doctoral thesis.

In the field of protein engineering, there are basically two ways of improving the biophysical properties of a given target protein. The first

relies on modifying the amino acid sequence of a protein by rational design. This approach usually requires a detailed structural or mechanistic description of the protein that is to be engineered. For many biochemical problems, however, either this information is not available, or, if it is available, the modifications required to rationally design a desired phenotype are too complex to be predicted. The second approach relies on engineering a protein by "irrational" design. This strategy is more powerful than rational design because it explores the evolutionary principles observed in nature, namely random mutagenesis and selection, to generate a desired molecular property. This concept is known as directed evolution and the underlying methods are called display methods or selection methods. The strength of directed evolution lies in the fact that a desired molecular property can be obtained based on very low input information. Moreover, because the methods employed by directed evolution explore the evolutionary potential of a protein in a combinatorial fashion, it is possible to evolve biophysical properties that are based on complex molecular interactions, such as the ones defining protein expression or thermodynamic stability.

The procedure for engineering the properties of a given protein involves two experimental steps. First, the introduction of modifications to the genetic sequence coding for a protein of interest, such as point mutations, insertions, or deletions, generates a set of diversified protein variants. As mentioned before, genetic diversity can be generated either by introducing specific mutations in a rational way or by mutagenizing the gene randomly. Second, the mutant protein variants are analyzed in the context of a functional assay in order to identify candidates showing the desired molecular property. For the second step, a distinction can be made between two experimental approaches for identifying improved candidates. If the number of mutant protein variants under investigation is relatively small – in the range of hundreds, as is often the case in rational design experiments – the different variants can be analyzed one-by-one by a *screen*. However, if a larger sequence space needs to be analyzed, for example in an attempt to evolve a particular protein variant based on a pool of millions of randomly mutagenized proteins, it is not possible to screen each variant individually. In this case the methods of directed evolution, which are based on genetic *selection* rather than screening, are more appropriate.

Although the principles of protein engineering have been applied to membrane proteins for more than 15 years[37,38], there have been far less studies on membrane proteins than on soluble proteins. Since appropriate methods for the directed evolution of membrane proteins were not

available until recently, most of the reported studies were based on screening small sets of mutants.

In 1999, Bowie and coworkers found that the thermal stability of the integral membrane protein diacylglycerol kinase could be increased by engineering cysteine-substituted variants[39]. Interestingly, two out of the 20 mutants they analyzed showed a stabilizing effect in detergent solution. In a subsequent study they screened a larger set of mutants and were able to further increase the thermal stability[40]. They concluded that stability-enhancing mutations appear to be relatively common in membrane proteins[41]. This observation inspired other research groups, including our own, to explore the potential of protein engineering for improving the biophysical properties of integral membrane proteins. Nordlund and coworkers for example have developed a dot-blot based method for screening relatively large sets of random mutants (in the range of thousands) for higher protein expression. Their approach, however, fails to distinguish between correctly folded and incorrectly folded protein. Tate and coworkers have applied a simple alanine scan to identify GPCR mutants showing increased thermal stability in detergent solution. This strategy has worked for a couple of GPCRs[42] and made possible the determination of the crystal structure of the β_1 adrenergic receptor[17]. Nevertheless, the strategy suffers form a basic limitation. Its application is limited to GPCRs that show a relatively high wild-type expression, which excludes the majority of GPCRs that fail to be expressed above a certain threshold level. Moreover, because the strategy is limited to the screening of individual point mutants it fails to explore the full evolutionary potential of a target protein.

Despite their limitations, the emerging protein engineering methods clearly open a new dimension to the field of membrane protein biochemistry and crystallography. The objective of the present thesis is the development and application of protein engineering methods that enable the improvement of the biophysical properties of integral membrane proteins, particularly of GPCRs. The main part of the studies is centered on the development of directed evolution methods for engineering highly expressed and stable GPCR variants.

1.4 Development of new methods for the directed evolution of GPCRs for expression and stability

The first study conducted during my doctoral thesis addressed the possibility of solving the experimental problems generally observed with GPCRs by a protein engineering approach that is focused on the biophysical properties of the protein surface. The idea was to convert the hydrophobic surface of a GPCR, which is naturally immersed in the phospholipid bilayer and causes most of the experimental trouble, into a hydrophilic surface. This would alleviate the problems caused by the hydrophobic nature of the protein and would allow to handle the resurfaced GPCR as if it was a soluble globular protein. The modifications on the protein surface should preserve the 7TM fold and the ligand binding competence of the GPCR. To this end, a strategy was developed for replacing the hydrophobic side chains on the protein surface, which normally have to be protected by detergent molecules, by hydrophilic side chains. We chose the β_2 adrenergic receptor as a model system.

The problem of engineering a water-soluble β_2 adrenergic receptor is far too complex to be solved by a rational design approach. We therefore decided to develop a combinatorial strategy that employs the principles of directed evolution, namely mutagenesis and selection. In **Chapter 2** of this thesis I present the theoretical concept and the synthesis on the DNA level of a semi-rational gene library that is designed for the selection of water-soluble β_2 adrenergic receptor variants. The genetic selection on this library, however, has not yet been carried out.

The complete resurfacing of a GPCR presented in **Chapter 2**, represents a very radical way of engineering a membrane protein and the risk of evolving a protein variant that shows impaired function is relatively high. A less radical approach for improving the biophysical properties of a GPCR is presented in **Chapter 3** of this thesis. It deals with the engineering of GPCRs by swapping specific domains, such as transmembrane helices, between homologous proteins. We produced receptor chimeras between the μ-opioid receptor (MOR), which behaved very poorly upon solubilization from membranes and the κ-opioid receptor, which proved somewhat more robust. Replacing a short stretch of sequence in MOR by the corresponding sequence from the KOR improved the behavior of the MOR to the level of that of the KOR while preserving the ligand specificity and affinity of the μ-opioid receptor.

The third and the fourth study (**Chapters 4 and 5**) in this thesis describe how the functional expression level and the biophysical stability of GPCRs can be strongly improved by introducing small changes to the amino acid sequence. These studies describe the development and critical assessment of a general method for the directed evolution of GPCRs. It represents the first report of applying the principles of directed evolution to the field of integral membrane proteins. The remainder of this introductory chapter is dedicated to a short overview on the application of directed evolution methods to the research field of integral membrane proteins. It leads to the introduction of the studies presented in the **Chapters 4 and 5**.

In the research field of soluble proteins, there exist a variety of directed evolution methods that exploit the natural principle of random mutagenesis and selection to isolate a desired mutant protein from a large pool of randomly mutagenized proteins. Each method is specified by a different mode of establishing a physical linkage between a genotype and its phenotype. As it is the case for living organisms, this physical linkage is absolutely necessary for the progression of an evolutionary cycle because the desired molecular property is selected on the protein level (phenotype) while the corresponding genetic information is encoded in the DNA (genotype). From a technical point of view, the identity of a particular selected protein is much easier to be revealed from its DNA sequence than from its amino acid sequence.

The different display methods for soluble proteins can be subdivided into two groups based on the mode by which the genotype-phenotype linkage is established. The *in-vivo* display methods make use of a living cell to establish the genotype-phenotype linkage and the protein is mostly displayed on the surface of the host cell. The most widely used host cell types for *cell-surface display* are bacteria and yeasts, but more complex cell types have been used as well. The *in-vitro* display methods, on the other hand, employ a cell-free and more direct mode for linking the protein to its corresponding DNA or mRNA. In *ribosome display* the protein is translated in vitro and the ribosome itself is used to link the translated protein with its mRNA[43-45]. In *phage display* the protein is encoded on the DNA encapsulated inside a phage particle and the protein is displayed on the phage surface[46,47]. In *mRNA display* the protein is translated in vitro and tethered directly to its mRNA[48-51]. In *in-vitro compartmentalization* both the protein and its mRNA are constrained within aqueous droplets (compartments), which are formed in a water-in-oil emulsion[45,52,53].

The cell-free in-vitro display methods developed for soluble proteins cannot be directly applied to membrane proteins, because the folding

mechanism of integral membrane proteins is not compatible with these methods. In most cases, integral membrane proteins fail to fold correctly under in-vitro conditions, which do not provide the physico-chemical environment of a lipid bilayer and the translocation machinery. However, the concepts of in-vivo display methods, such as cell-surface display, could potentially be adapted for the directed evolution of integral membrane proteins. **Chapter 4** of the thesis describes the development of an efficient in-vivo display method for the directed evolution of GPCRs in *Escherichia coli*. It aims at improving functional expression, biophysical stability, and ligand selectivity. The performance of the method was critically assessed using the model protein neurotensin receptor NTR1. In a follow-up study presented in **Chapter 5**, I further validated the method by applying it successfully to three additional human GPCRs, including a GPCR that can hardly be expressed in functional form in *E. coli*.

1.5 References:

1. http://www.gpcr.org/7tm/
2. http://www.iuphar-db.org/
3. Bjarnadottir, T.K., Gloriam, D.E., Hellstrand, S.H., Kristiansson, H., Fredriksson, R. & Schioth, H.B. **Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse**. *Genomics* **88**, 263-273 (2006).
4. Foord, S.M., Bonner, T.I., Neubig, R.R., Rosser, E.M., Pin, J.P., Davenport, A.P., Spedding, M. & Harmar, A.J. **International Union of Pharmacology. XLVI. G protein-coupled receptor list**. *Pharmacol Rev* **57**, 279-288 (2005).
5. Schioth, H.B. & Fredriksson, R. **The GRAFS classification system of G-protein coupled receptors in comparative perspective**. *Gen Comp Endocrinol* **142**, 94-101 (2005).
6. Attwood, T.K. & Findlay, J.B. **Fingerprinting G-protein-coupled receptors**. *Protein Eng* **7**, 195-203 (1994).
7. Kolakowski, L.F., Jr. **GCRDb: a G-protein-coupled receptor database**. *Receptors Channels* **2**, 1-7 (1994).
8. Rodbell, M. **The role of hormone receptors and GTP-regulatory proteins in membrane transduction**. *Nature* **284**, 17-22 (1980).
9. Lefkowitz, R.J. & Shenoy, S.K. **Transduction of receptor signals by beta-arrestins**. *Science* **308**, 512-517 (2005).
10. Heuss, C. & Gerber, U. **G-protein-independent signaling by G-protein-coupled receptors**. *Trends Neurosci* **23**, 469-475 (2000).
11. Overington, J.P., Al-Lazikani, B. & Hopkins, A.L. **How many drug targets are there?** *Nat Rev Drug Discov* **5**, 993-996 (2006).
12. http://blanco.biomol.uci.edu/
13. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M. & Miyano, M. **Crystal structure of rhodopsin: A G protein-coupled receptor**. *Science* **289**, 739-745 (2000).
14. Murakami, M. & Kouyama, T. **Crystal structure of squid rhodopsin**. *Nature* **453**, 363-367 (2008).
15. Rasmussen, S.G., Choi, H.J., Rosenbaum, D.M., Kobilka, T.S., Thian, F.S., Edwards, P.C., Burghammer, M., Ratnala, V.R., Sanishvili, R., Fischetti, R.F., Schertler, G.F., Weis, W.I. & Kobilka, B.K. **Crystal structure of the human beta2 adrenergic G-protein-coupled receptor**. *Nature* **450**, 383-387 (2007).
16. Cherezov, V., Rosenbaum, D.M., Hanson, M.A., Rasmussen, S.G., Thian, F.S., Kobilka, T.S., Choi, H.J., Kuhn, P., Weis, W.I., Kobilka, B.K. & Stevens, R.C. **High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor**. *Science* **318**, 1258-1265 (2007).
17. Warne, T., Serrano-Vega, M.J., Baker, J.G., Moukhametzianov, R., Edwards, P.C., Henderson, R., Leslie, A.G., Tate, C.G. & Schertler, G.F. **Structure of a beta1-adrenergic G-protein-coupled receptor**. *Nature* **454**, 486-491 (2008).
18. Jaakola, V.P., Griffith, M.T., Hanson, M.A., Cherezov, V., Chien, E.Y., Lane, J.R., Ijzerman, A.P. & Stevens, R.C. **The 2.6 Ångstrom crystal structure of a human A2A adenosine receptor bound to an antagonist**. *Science* **322**, 1211-1217 (2008).

19. Ratnala, V.R., Swarts, H.G., VanOostrum, J., Leurs, R., DeGroot, H.J., Bakker, R.A. & DeGrip, W.J. **Large-scale overproduction, functional purification and ligand affinities of the His-tagged human histamine H1 receptor**. *Eur J Biochem* **271**, 2636-2646 (2004).
20. Krepkiy, D., Gawrisch, K. & Yeliseev, A. **Expression and purification of CB2 for NMR studies in micellar solution**. *Protein Pept Lett* **14**, 1031-1037 (2007).
21. White, J.F., Trinh, L.B., Shiloach, J. & Grisshammer, R. **Automated large-scale purification of a G protein-coupled receptor for neurotensin**. *FEBS Lett* **564**, 289-293 (2004).
22. Kobilka, B.K. & Deupi, X. **Conformational complexity of G-protein-coupled receptors**. *Trends Pharmacol Sci* **28**, 397-406 (2007).
23. Ostermeier, C., Iwata, S., Ludwig, B. & Michel, H. **Fv fragment-mediated crystallization of the membrane protein bacterial cytochrome c oxidase**. *Nat Struct Biol* **2**, 842-846 (1995).
24. Hunte, C., Koepke, J., Lange, C., Rossmanith, T. & Michel, H. **Structure at 2.3 Å resolution of the cytochrome bc(1) complex from the yeast Saccharomyces cerevisiae co-crystallized with an antibody Fv fragment**. *Structure* **8**, 669-684 (2000).
25. Dutzler, R., Campbell, E.B. & MacKinnon, R. **Gating the selectivity filter in ClC chloride channels**. *Science* **300**, 108-112 (2003).
26. Sennhauser, G., Amstutz, P., Briand, C., Storchenegger, O. & Grütter, M.G. **Drug export pathway of multidrug exporter AcrB revealed by DARPin inhibitors**. *PLoS Biol* **5**, e7 (2007).
27. Link, A.J., Skretas, G., Strauch, E.M., Chari, N.S. & Georgiou, G. **Efficient production of membrane-integrated and detergent-soluble G protein-coupled receptors in Escherichia coli**. *Protein Sci* **17**, 1857-1863 (2008).
28. Skretas, G. & Georgiou, G. **Genetic analysis of G protein-coupled receptor expression in Escherichia coli: inhibitory role of DnaJ on the membrane integration of the human central cannabinoid receptor**. *Biotechnol Bioeng* **102**, 357-367 (2009).
29. Klammt, C., Schwarz, D., Eifler, N., Engel, A., Piehler, J., Haase, W., Hahn, S., Dotsch, V. & Bernhard, F. **Cell-free production of G protein-coupled receptors for functional and structural studies**. *J Struct Biol* **158**, 482-493 (2007).
30. Leitz, A.J., Bayburt, T.H., Barnakov, A.N., Springer, B.A. & Sligar, S.G. **Functional reconstitution of Beta2-adrenergic receptors utilizing self-assembling Nanodisc technology**. *Biotechniques* **40**, 601-602, 604, 606, passim (2006).
31. Borch, J. & Hamann, T. **The nanodisc: a novel tool for membrane protein studies**. *Biol Chem* **390**, 805-814 (2009).
32. Berger, B.W., Garcia, R.Y., Lenhoff, A.M., Kaler, E.W. & Robinson, C.R. **Relating surfactant properties to activity and solubilization of the human adenosine a3 receptor**. *Biophys J* **89**, 452-464 (2005).
33. Newstead, S., Ferrandon, S. & Iwata, S. **Rationalizing alpha-helical membrane protein crystallization**. *Protein Sci* **17**, 466-472 (2008).
34. Drew, D.E., von Heijne, G., Nordlund, P. & de Gier, J.W. **Green fluorescent protein as an indicator to monitor membrane protein overexpression in Escherichia coli**. *FEBS Lett* **507**, 220-224 (2001).

35. Hammon, J., Palanivelu, D.V., Chen, J., Patel, C. & Minor, D.L., Jr. **A green fluorescent protein screen for identification of well-expressed membrane proteins from a cohort of extremophilic organisms**. *Protein Sci* **18**, 121-133 (2009).
36. Geertsma, E.R., Groeneveld, M., Slotboom, D.J. & Poolman, B. **Quality control of overexpressed membrane proteins**. *Proc Natl Acad Sci U S A* **105**, 5722-5727 (2008).
37. Mingarro, I., von Heijne, G. & Whitley, P. **Membrane-protein engineering**. *Trends Biotechnol* **15**, 432-437 (1997).
38. Popot, J.L. & Saraste, M. **Engineering membrane proteins**. *Curr Opin Biotechnol* **6**, 394-402 (1995).
39. Lau, F.W., Nauli, S., Zhou, Y. & Bowie, J.U. **Changing single side-chains can greatly enhance the resistance of a membrane protein to irreversible inactivation**. *J Mol Biol* **290**, 559-564 (1999).
40. Zhou, Y. & Bowie, J.U. **Building a thermostable membrane protein**. *J Biol Chem* **275**, 6975-6979 (2000).
41. Bowie, J.U. **Stabilizing membrane proteins**. *Curr Opin Struct Biol* **11**, 397-402 (2001).
42. Tate, C.G. & Schertler, G.F. **Engineering G protein-coupled receptors to facilitate their structure determination**. *Curr Opin Struct Biol* **19**, 386-395 (2009).
43. Mattheakis, L.C., Bhatt, R.R. & Dower, W.J. **An in vitro polysome display system for identifying ligands from very large peptide libraries**. *Proc Natl Acad Sci U S A* **91**, 9022-9026 (1994).
44. Hanes, J. & Plückthun, A. **In vitro selection and evolution of functional proteins by using ribosome display**. *Proc Natl Acad Sci U S A* **94**, 4937-4942 (1997).
45. He, M. & Taussig, M.J. **Antibody-ribosome-mRNA (ARM) complexes as efficient selection particles for in vitro display and evolution of antibody combining sites**. *Nucleic Acids Res* **25**, 5132-5134 (1997).
46. Smith, G.P. **Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface**. *Science* **228**, 1315-1317 (1985).
47. Dunn, I.S. **Phage display of proteins**. *Curr Opin Biotechnol* **7**, 547-553 (1996).
48. Roberts, R.W. & Szostak, J.W. **RNA-peptide fusions for the in vitro selection of peptides and proteins**. *Proc Natl Acad Sci U S A* **94**, 12297-12302 (1997).
49. Nemoto, N., Miyamoto-Sato, E., Husimi, Y. & Yanagawa, H. **In vitro virus: bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro**. *FEBS Lett* **414**, 405-408 (1997).
50. Kurz, M., Gu, K., Al-Gawari, A. & Lohse, P.A. **cDNA - protein fusions: covalent protein - gene conjugates for the in vitro selection of peptides and proteins**. *ChemBioChem* **2**, 666-672 (2001).
51. Liu, R., Barrick, J.E., Szostak, J.W. & Roberts, R.W. **Optimized synthesis of RNA-protein fusions for in vitro protein selection**. *Methods Enzymol* **318**, 268-293 (2000).

52. Doi, N. & Yanagawa, H. **STABLE: protein-DNA fusion system for screening of combinatorial protein libraries in vitro**. *FEBS Lett* **457**, 227-230 (1999).
53. Tawfik, D.S. & Griffiths, A.D. **Man-made cell-like compartments for molecular evolution**. *Nat Biotechnol* **16**, 652-656 (1998).

Chapter 2

Conversion of the β_2 adrenergic receptor into a water-soluble analog

2.1 Abstract

Despite enormous efforts made in generating atomic-resolution structures of G-protein coupled receptors, there are only five structures solved so far. Here we propose a radically new strategy for overcoming the experimental problems inherent to the study of integral membrane proteins by engineering water-soluble analogs of GPCRs suitable for X-ray crystallography. We use rational design for the generation of gene libraries from which improved molecules can be selected and further evolved towards the desired biophysical and functional properties. Starting from a homology model of the β_2 adrenergic receptor structure, computational methods are used to identify amino acid positions suitable for randomization while limiting the combinatorial complexity of the library to a level that can be handled by modern in-vitro selection technologies. Ribosome display will allow the selection and further improvement of receptor variants capable of achieving their native fold and ligand binding competence in an aqueous environment.

2.2 Introduction

Membrane and water-soluble proteins both show similarly organized protein cores, predominantly composed of closely packed hydrophobic amino acids. The main difference between soluble proteins and integral membrane proteins lies in the fact that the molecular surface of water-soluble proteins is predominantly composed of hydrophilic amino acids, while in membrane proteins, those parts of the molecular surface that are normally immersed in the phospholipid bilayer are predominantly hydrophobic. In water-soluble globular proteins, the hydropathic contrast between the hydrophobic protein interior and the hydrophilic surface plays a major role in the folding process. Increasing the proportion of solvent-exposed hydrophobic amino acids in such proteins can significantly lower their folding efficiency. In contrast, membrane proteins utilize the secretion/membrane insertion machinery of the cell to insert the protein into the membrane, where the hydrophobic surface residues are surrounded and chaperoned by lipids.

Here we propose to replace the phospholipid-exposed hydrophobic residues within the transmembrane (TM) domains of the GPCRs with more hydrophilic residues to engineer water-soluble variants of GPCRs capable of folding in aqueous solutions. The proof-of-concept for the idea that membrane proteins can be resurfaced and converted into soluble globular proteins while maintaining a similarly packed protein core has been presented[1-4]: Frank and coworkers were the first to design resurfaced mutants of the single transmembrane helix protein phospholamban. These variants showed increased solubility over the wild-type protein and associated into a native-like pentameric coiled-coil structure. The crystal structure of a water-soluble phospholamban analog could be solved[5]. More recently, a computational design procedure was applied to engineer water-soluble variants of the prokaryotic potassium channel KcsA[4]. The resulting analogs were expressed with high yields in *E. coli* and mimicked some of the functional properties of the wild-type channel in aqueous solution, such as oligomerization state and binding of specific channel- blocking agents. However, efforts to engineer water-soluble resurfaced mutants of the larger protein bacteriorhodopsin resulted in proteins that were unable to adopt the wild-type fold and as a consequence showed very low solubility in water[6]. The positive results of successful solubilization strategies in other groups[1-5] inspired us to develop new methods for redesigning the transmembrane domain of GPCRs. This study aims to produce GPCR analogs, which are soluble and stable in water and can be expressed in high quantities to enable atomic-resolution biophysical studies. In this chapter I present the

theoretical concept and experimental synthesis of a semi-rational combinatorial gene library that is designed for the selection of water-soluble analogs of the β_2 adrenergic receptor (ADRB2).

2.3 Results and discussion

2.3.1 Identifying phospholipid-exposed amino acids for randomization

The problem of engineering a water-soluble β_2 adrenergic receptor analog is far too complex to be solved by a rational design approach[6]. We therefore decided to develop a combinatorial strategy that employs the concept of directed evolution involving the mutagenesis of the β_2 adrenergic receptor gene (ADRB2) and the isolation of water-soluble receptor analogs by in-vitro selection techniques such as ribosome display. The central idea of our approach is the computational design of a combinatorial gene library that contains receptor variants that fold into the 7TM fold in aqueous solution. These water-soluble analogs can then be isolated from the library via ribosome display by selecting for their ability to bind specific ADRB2 ligands, such as the antagonist alprenolol. In this chapter I describe the conceptual design and the synthesis of the gene library. The selection by ribosome display has not been carried out yet.

In order to convert the hydrophobic surface of ADRB2 into a hydrophilic one, we first identified the highly surface-exposed amino acid positions that normally are immersed in the phospholipid bilayer. These positions would then be targeted for randomization in the gene library. To identify these positions we followed two lines of evidence. First, we constructed and analyzed a structural homology model for ADRB2. Second, we analyzed multiple sequence alignments of the adrenergic receptor family to identify highly variable – and thus presumably solvent-exposed – amino acid positions.

When we started this project in 2004, there was no high-resolution crystal structure available for the β_2 adrenergic receptor, which made it difficult to identify the appropriate positions for randomization. We therefore constructed a homology model of the receptor based on the crystal structure of rhodopsin (PDB code 1f88), which was the only atomic-resolution structure of a GPCR available at that time[7]. This structure represents the non-signaling state of rhodopsin. Because we were mainly interested in the spacial orientation of the surface-exposed, lipid-contacting residues, rather than a detailed model of the entire receptor structure, we limited the homology model to the region of the seven TM helices.

ADRB2 and rhodopsin both belong to the rhodopsin-like family of GPCRs. Although they share very low sequence identity (about 20% in the 7TM region), they do share some highly conserved sequence

characteristics (GPCR signature motifs) that are found in each of the seven TM helices. The homology model building started with the pairwise sequence alignment of human ADRB2 and bovine rhodopsin (**Fig. 2.1a**). The conserved signature motifs served as anchoring points for aligning the seven TM helices correctly.

The initial structural model of ADRB2 was built by assigning the backbone coordinates for each aligned amino acid. The helix ends of ADRB2 were defined based on the rhodopsin structure. The ADRB2 side chains were then assigned to the backbone. This initial model is a simple duplication of the rhodopsin backbone carrying the ADRB2 side chains.

The model was then refined. The spacial orientations of all side chains in the model were optimized by sampling a backbone dependent rotamer library. The orientations of residues comprising the GPCR signature motifs were kept as in the rhodopsin structure. All other side chains were optimized to remove steric clashes, and saturate hydrogen-bonding contacts. After optimizing the orientations of all side chains, strong steric clashes remained near the extracellular end of TM2 and along the interface between TM3 and TM4. These clashes could be removed by minor adjustments of the helix orientations of TM2 and TM4. TM2 was straightened up by moving the extracellular part of the helix away from the protein core. The helix kink in TM2 served as a hinge for this movement. Increasing the distance between TM3 and TM4 eliminated the steric clash between these two helices. To this end, the entire TM4, which is highly exposed to phospholipids, was moved away from TM3. The strong helical distortions (kinks induced by π- and 3_{10} helical elements) observed in the template structure of rhodopsin were also incorporated into the homology model. The superposition of the final ADRB2 model on the ADRB2 crystal structure solved in 2007[8] (PDB code 2rh1) is illustrated in **Figure 2.1b**. The superposition shows a low RMSD value of 1.1 Å between the Cα atoms of the model and the ADRB2 crystal structure.

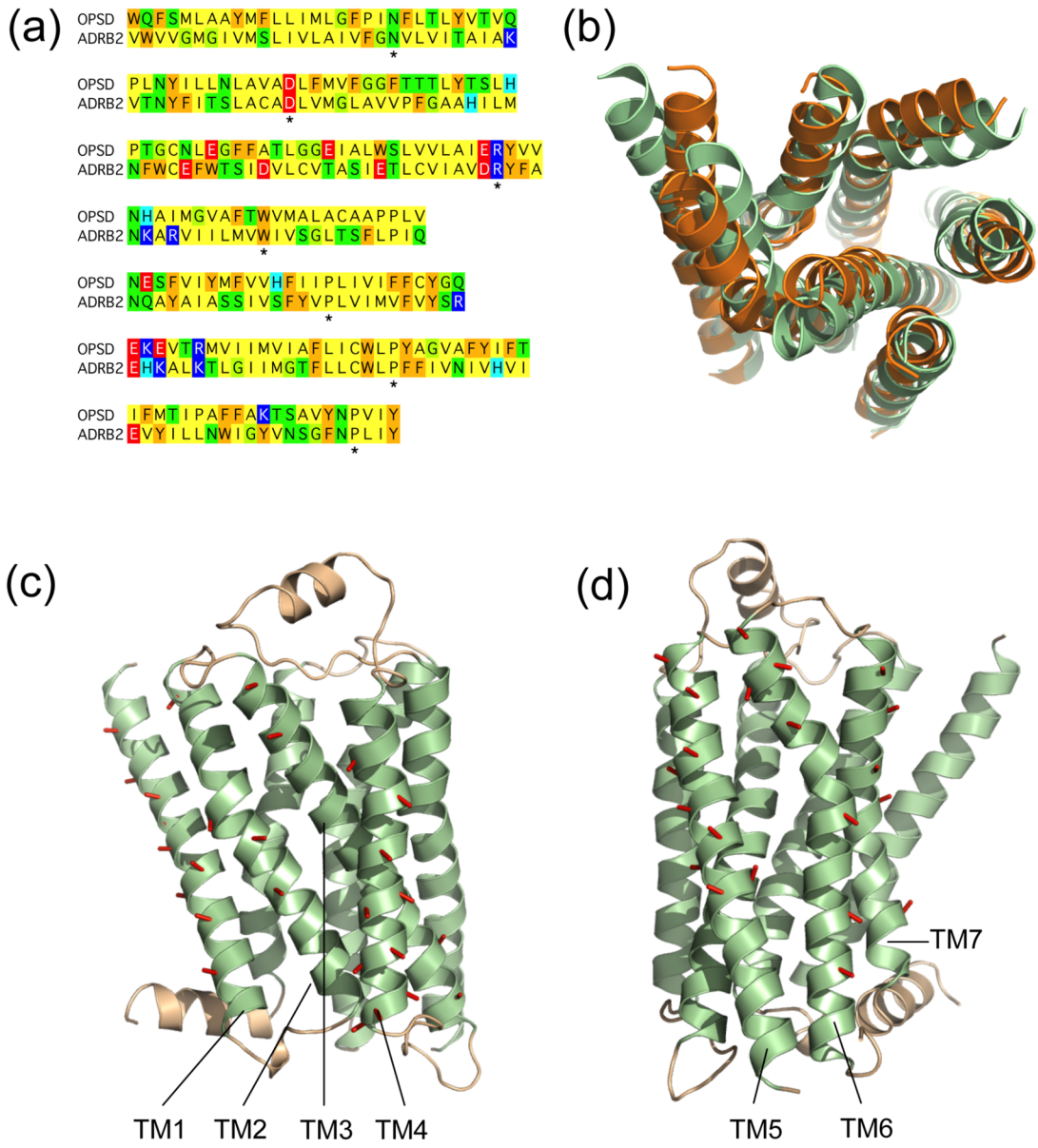


Figure 2.1: Homology modeling of the human ADRB2. **(a)** Pairwise sequence alignment of the seven TM helices of bovine rhodopsin and ADRB2 (TM1 through TM7 from top to bottom). The signature residue that is most conserved in each TM among the rhodopsin family of GPCRs is indicated (*). **(b)** Superposition of the Cα backbone coordinates of the ADRB2 homology model (orange ribbon) and the ADRB2 crystal structure (green ribbon, PDB code 2rh1). **(c, d)** The 54 randomized amino acid positions placed on the ADRB2 crystal structure as red Cα-Cβ vectors.

The quality of the refined homology model was verified by its consistency with experimental data. We were most interested to verify that the seven helices were oriented correctly to each other, i.e. that the phospholipid-exposed faces of each helix were indeed positioned on the

receptor surface. To this end, we reanalyzed data from an earlier study on the dopamine receptor D_2, a close homolog of ADRB2, in the context our homology model. In this study, Javitch and coworkers[9] had used the substituted-cysteine accessibility method (SCAM) for mapping the surface of the ligand-binding crevice in the D_2 receptor. They mutated every residue of the seven TM helices to Cys, one at a time, and probed its accessibility to a charged sulfhydryl reagent. Water-accessible Cys residues in the ligand-binding crevice are expected to be vastly more reactive to the reagent than Cys facing lipid or facing the tightly packed protein core. By mapping their data on our homology model of ADRB2 we found that the orientations of all seven helices were in good agreement with the experimentally found helix orientations in the D_2 receptor. The homology model ADRB2 thus allowed identifying the most phospholipid-exposed amino acid positions on the surface of each of the seven TM helices.

In addition to the structural information obtained from the homology model, we investigated a second line of evidence to identify appropriate amino acid positions for randomization. We constructed a multiple sequence alignment of nearly hundred β_2 adrenergic receptor homologs from different species retrieved from NCBI's GenBank and analyzed the amino acid variability for each aligned position. We supposed that phospholipid-exposed amino acid positions would be less conserved than positions facing the protein core. The positions showing the highest variability were mostly populated by the hydrophobic amino acids Ala, Gly, Ile, Leu, Val or Phe. In agreement with the structural analysis of the homology model, we found that the variable positions mostly overlapped with the phospholipid-exposed positions of the ADRB2 model (**Fig. 2.2**).

Based on the analysis of the homology model and of the sequence variability data, we identified 54 amino acid positions that we assumed were highly phospholipid-exposed, and we decided to randomize those positions in the gene library of ADRB2. The 54 randomized positions correspond to 28% of the all amino acids in the seven TM helices and 12% of the full-length ADRB2 receptor. When the crystal structure of the human β_2 adrenergic receptor was published in 2007[8,10] we mapped the 54 positions on this structure and found that all of them were indeed surface exposed as illustrated in **Figures 2.1c** and **2.1d**.

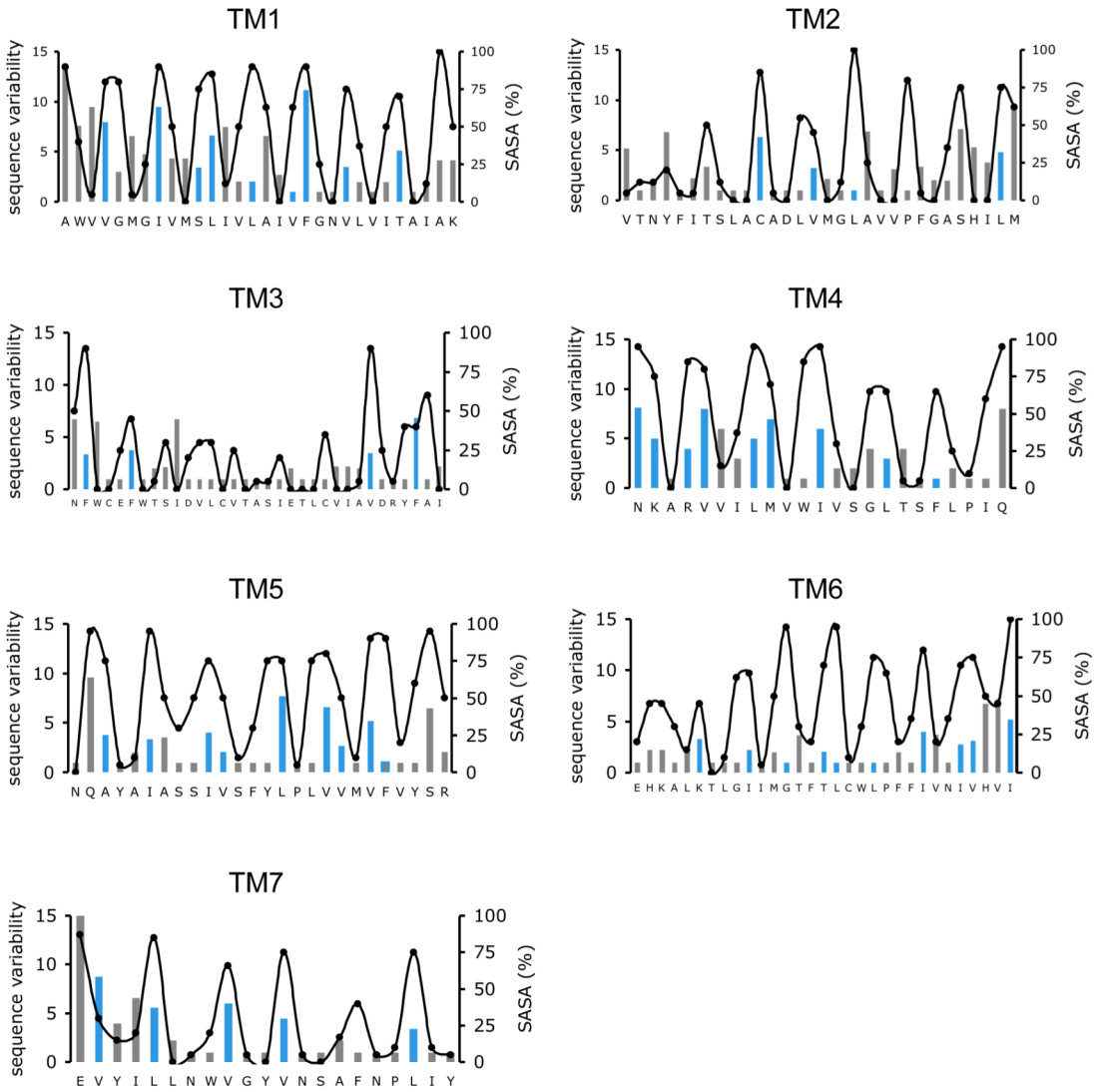


Figure 2.2: Correlation between sequence variability of 95 aligned ADRB2 homologs of different species (bars) and solvent accessible surface area in the ADRB2 homology model (•, SASA) for every residue in the seven TM helices. These data, along with visual inspection of the ADRB2 homology model, were used to define the 54 phospholipid-exposed amino acid positions for randomization (blue bars). The sequence variability at a given amino acid position is calculated by dividing the total number of different amino acids at that position by the frequency of the most common amino acid at that position.

2.3.2 Design of a hydrophilic ADRB2 surface

To design the water-soluble surface of ADRB2 based on the 54 phospholipid-exposed positions, we used a subset to the amino acids EKRQAVLF to replace the wild-type amino acids (colored in blue in **Fig. 2.3**). The hydrophilic and charged amino acids (EKRQ) were used to confer a water-soluble character to the receptor. The hydrophobic amino acids (AVLF) were used to preserve of the wild-type-like hydrophobic character of a given position in case a hydrophilic amino acid would not be tolerated at this position.

Not all 54 positions were randomized to the same extent. Rather we defined an optimal subset of replacement amino acids for each randomized position, taking into consideration the wild-type amino acid at a given position. For example, Leu is the wild-type amino acid at position 45 in TM1 of ADRB2, and we incorporated the amino acid subset EKRQAL at this position, which includes the wild-type Leu. Designing a restricted subset of amino acids at each position has the advantage of strongly reducing the theoretical complexity of the library. **Figure 2.3** shows the design of the 54 positions in the seven TM helices (blue boxes)

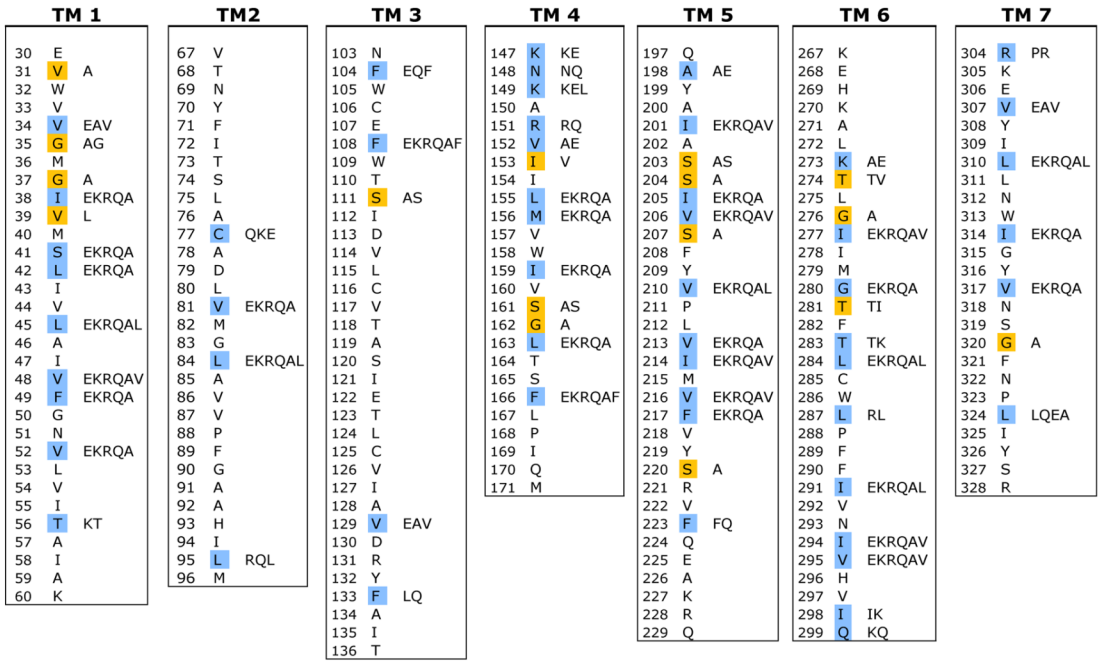


Figure 2.3: Design of the seven TM helices. The 54 amino acid positions that are randomized for increasing the water-soluble surface of ADRB2 (blue) and the 16 positions that are randomized or replaced to increase the overall α-helical content of a helix (yellow) are shown. The amino acid types that are inserted at a given position in the library are shown to the right of each helix.

The redesign of each TM helix of ADRB2 resulted in seven highly randomized helices; each of them showed a theoretical complexity of up to 10^6 different sequence variants. We wondered how likely the redesigned helices would adopt α-helical secondary structure in aqueous solution. We therefore took an abstract look at each helix by analyzing their characteristics as independently folding peptide chains – decoupled from the structural context of the 7TM fold. To this end, we employed a computational model – AGADIR – that predicts the α-helix formation for peptides in aqueous solution with high accuracy[11,12]. The AGADIR model provides a quantitative representation of experimentally derived statistical parameters, such as intrinsic secondary structure propensities and position dependence of the 20 amino acids, as well as physical parameters, such as electrostatic or side-chain side-chain interactions, that defines the folding of α-helices.

We applied the AGADIR algorithm to the each of the seven helical peptide ensembles to predict their tendency of forming α-helical secondary structure. The algorithm computed the helical content (at the single residue level) for every possible peptide sequence that would theoretically exist in a helix sub-library (**Fig. 2.3**). The results for the peptide sequences corresponding to the seven wild-type TM helices of ADRB2, which show a very strong hydrophobic character, showed a very low α-helical secondary structure content of 1-2% (averaged over the entire peptide). The analysis of the redesigned helices, however, showed two interesting results. First, the redesigned helices were very different in their α-helical content. In the libraries of TM1 and TM4 we found many sequence variants that showed more than 25% helical content on average. In contrast, most of the sequences in the libraries of TM2, TM3, TM5, TM6 and TM7 showed less than 5% α-helical content. This difference is most probably due to the fact that TM1 and TM4 represent rather regularly shaped helices in the ADRB2 crystal structure, while TM2, TM5, TM6 and TM7 are distorted by proline-induced kinks. The low α-helical content of TM3, however, most probably relates to its strong hydrophobic character, which it retains in the redesigned version of the receptor because TM3 is largely buried in the receptor core. A second finding from the analysis by AGADIR was that replacing some of the wild-type Gly, Ser or Val residues by Ala or Leu strongly increased the α-helical content of TM1, TM4 and TM5. For example, replacing Gly37, Val39 and Ser41 by Ala or Leu in TM1 led to a large increase in α-helical content as illustrated in **Figure 2.4**. Gly50, however, which is predicted to break off TM1, could not be replaced, because it is highly conserved in the GPCR superfamily.

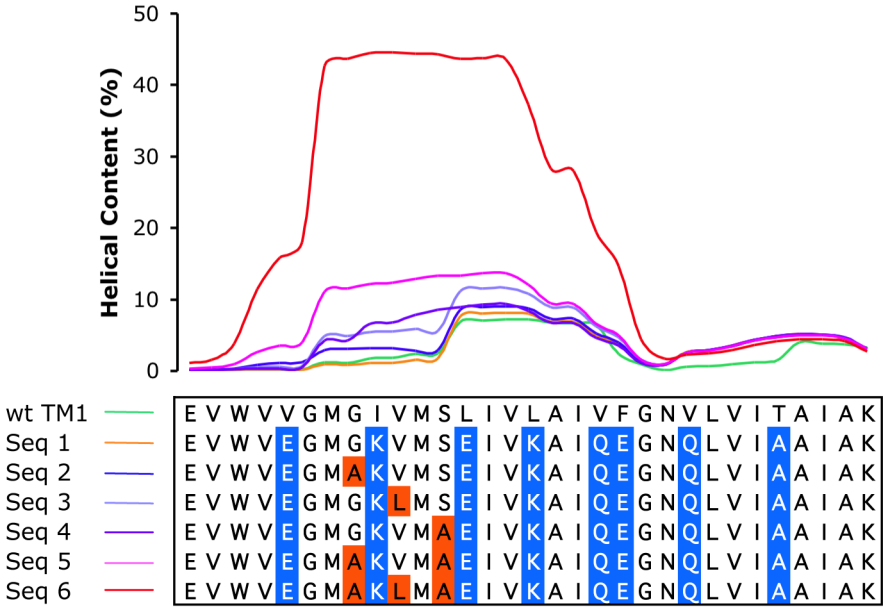


Figure 2.4: Visualization of a calculation by AGADIR on seven designed sequences of TM1. The percentage helical content for each amino acid position is shown on top, the corresponding amino acid peptides of TM1 at the bottom. Replacing Gly, Val or Ser residues by Ala or Leu (orange boxes) leads to a strong increase of the α-helical content of TM1.

The results obtained by the AGADIR calculations suggested that the helical content of the library could be increased by small modifications to the sequence. We therefore decided to include 16 positions in the redesign of the receptor (colored in yellow in **Fig. 2.3**) to improve the overall helical content of the library (in addition to the 54 positions, which were randomized with a subset of the amino acids EKRQAVLF to confer water-solubility to the receptor). Many of the modifications suggested by the AGADIR calculations were in good agreement with data obtained from multiple sequence alignments of the adrenergic receptor family, where the suggested amino acids also occurred at homologous positions of closely related adrenergic receptors.

As a last modification in the library, we decided to mutate five Cys residues in the C-terminus (at the positions 327, 341, 354, 405, and 433) to Ala or Ser in order to limit the formation of non-specific disulfide bonds during protein expression. The Cys replacements were shown to be accommodated in the ADRB2 model structure without interfering with receptor function. The final redesign of the seven TM helices is summarized in **Figure 2.3**.

2.3.3 Synthesis of DNA encoding the combinatorial ADRB2 library

The full-length cDNA of the human ADRB2 consists of 1320 nucleotide base pairs that encode 440 amino acids; the library design demands a randomization or modification of 70 amino acid positions (54 + 16). The final combinatorial gene library therefore consists of seven highly randomized sequence segments (the TM helices), which are connected by six wild-type-like sequence segments (the intracellular and extracellular loops and the N- and C-termini). The synthesis of this complex library on the nucleotide level represents a difficult genetic engineering problem. We decided to incorporate the randomized positions into the cDNA by a complete de novo resynthesis of the entire receptor cDNA. The cloning strategy is based on dissecting the 1320 base pairs into 12 building blocks; each block is synthesized individually and the 12 blocks are then assembled to form the full-length library (assembly scheme in **Fig. 2.5**).

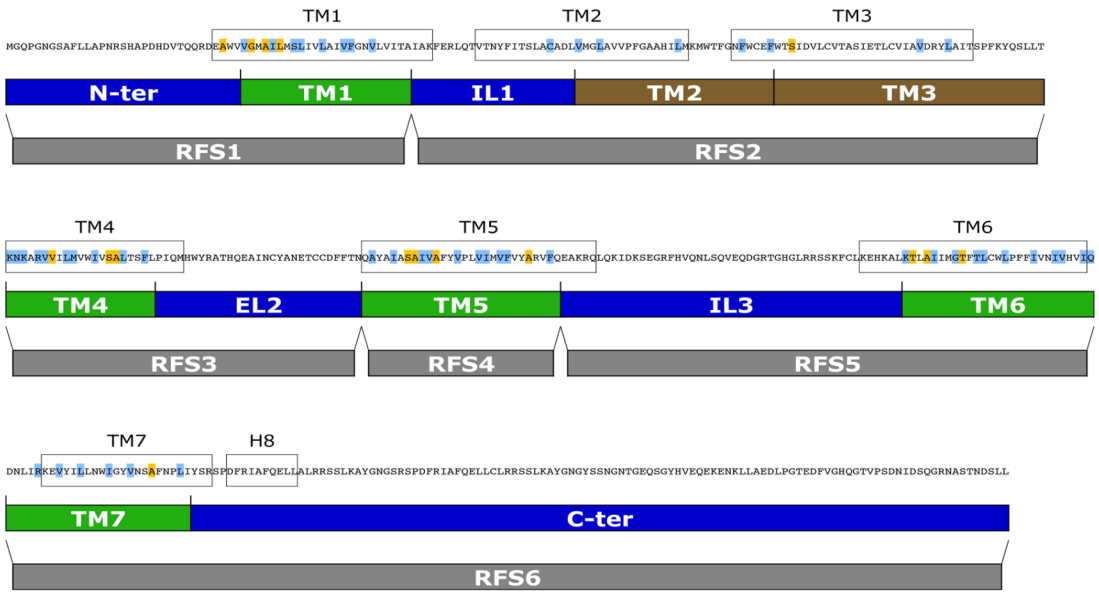


Figure 2.5: Schematic representation of the assembly strategy of the gene library for ADRB2 by 12 building blocks. The amino acid sequence of ADRB2 and the randomized positions are indicated on top of the 12 blocks. The full-length library was assembled as follows: First, the DNA for the 12 building blocks was synthesized (gree, brown and blue bars). The blocks for the highly randomized helices (green bars) were synthesized by restriction and ligation of four DNA fragments per block (see **Fig. 2.6**). The building blocks for the moderately randomized helices (brown bars) were synthesized by assembly PCR using overlapping oligonucleotides. The building blocks that encode only wt DNA (dark blue bars), were synthesized by a commercial vendor. The 12 building blocks were then ligated to form six modules for reading-frame selection (grey bars). The six modules can be ligated to form the full-length library of ADRB2.

The synthesis of the highly randomized building blocks that encode the TM helices 1, 4, 5, 6 and 7 (green bars in **Fig. 2.5**) presented the biggest conceptual and experimental challenge. These blocks are composed of about 20 to 30 triplet codons (60 to 90 base pairs) of which every second to third codon is highly randomized (encoding a subset of the amino acids EKRQAVLF). Unfortunately, it is not possible to assemble such helical building blocks (80 to 110 nucleotides long, including the terminal restriction sites) from one single oligonucleotide without facing serious problems regarding the oligonucleotide quality (mainly problems of oligonucleotide shortening). We therefore decided to synthesize each randomized helical building block from four shorter fragments, which were encoded on four randomized oligonucleotides (**Fig. 2.6**). The middle part of each oligonucleotide encodes five to six amino acids positions of the helix (colored in grey in **Fig. 2.6**). It is flanked on both sides by a constant stretch of nucleotides that encodes a recognition site for a type IIs restriction endonucleases (colored in green in **Fig. 2.6**). This oligonucleotide design allows the precise enzymatic cleavage of each of the four DNA fragments on both sides (fragments a, b, c and d in **Fig. 2.6**). By this cleavage, the flanking parts of the DNA fragment are removed and specific 5'-overhangs are formed on the middle part that encodes a fragment of the helix. The four restricted fragments are then connected by enzymatic ligation to form a complete helical building block. The quality of the synthesized helical building blocks was verified by the sequencing of 19 randomly picked sequences of TM1 and TM5. More than 50% of the sequences showed the expected sequence pattern and reading frame.

The remaining seven building blocks (dark blue bars and brown bars in **Fig. 2.5**) account for about 70% of the receptor sequence and they encode either few randomized positions (as in the building blocks TM2 and TM3) or no randomized positions at all (as in IL1, EL2, IL3 and the N- and C-termini). To obtain the DNA for these building blocks, we employed a simpler synthesis strategy than for the highly randomized building blocks for the TM helices. The building blocks for TM2 and TM3 were constructed by assembly PCR using overlapping oligonucleotides. We verified the quality of the assembled DNA by the sequencing of eight randomly picked sequences and found that more than 50% of the sequences showed the expected sequence pattern and reading frame. The remaining five building blocks, which encode wild-type DNA only (dark blue bars in **Fig. 2.5**), were ordered in a double-stranded DNA format from a commercial vendor.

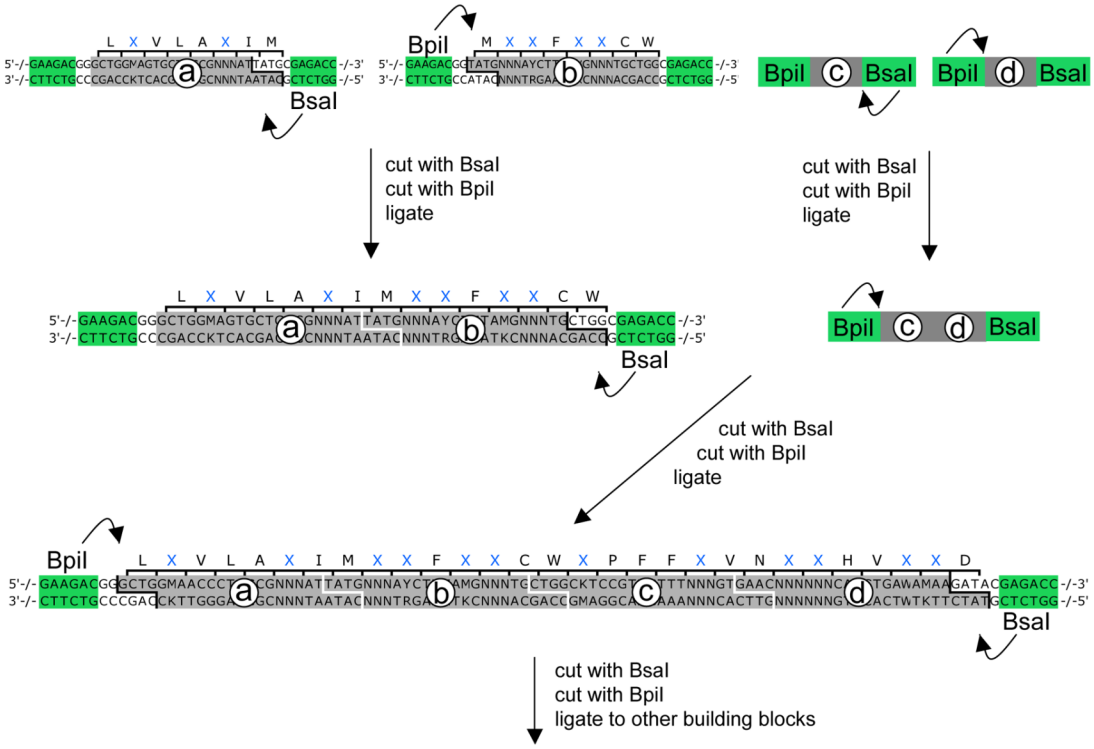


Figure 2.6: Synthesis at the DNA level of a building block encoding a highly randomized helix. One building block (e.g. TM6 shown here) is synthesized from four double-stranded DNA fragments denoted **a**,**b**,**c** and **d**. The two fragments **a** and **b** are shown at the DNA level on the top left, the two other fragments **c** and **d** are shown schematically on the top right. Each fragment encodes in its middle part a quarter of the nucleotide base pairs of a designed TM helix (colored in grey). Restriction enzyme recognition sites for BpiI and BsaI (colored in green) flank the middle part. Note that the recognition sites for BpiI and BsaI are distant from their cleavage sites (continuous line in the DNA sequence). The amino acid sequences are indicated above the genetic sequences (X indicates a randomized amino acid position). The synthesis of a complete building block works as follows. First, the four fragments are cleaved by either BpiI or BsaI. Fragment **a** is then ligated to fragment **b**; fragment **c** to **d**. In the second step, the two resulting fragments are again cleaved by either BpiI or BsaI. They are then ligated to form the full-length helical building block composed of the grey coding regions **a**, **b**, **c** and **d**. The resulting building block can then be ligated to its neighboring building blocks to form the full-length library (schematically shown in **Fig. 2.5**).

As mentioned above, the sequence analysis of the highly randomized helical building blocks had revealed that about 50% of the sequences were not in the correct reading frame due to insertions or deletions in the nucleotide sequence. While this error rate may be considered as relatively small on the level of a single building block, it would strongly increase when the full-length library is assembled from the individual building

blocks. Based on the 50% error rate, the fraction of correct full-length sequences in the final library was estimated to be only about 1% (0.5^6 = 0.012). This represented a 100-fold reduction in the sequence space that we intended to sample in the subsequent selection by ribosome display. In order to improve the quality of the library before starting the selection by ribosome display we decided to perform an experiment that aims at removing the faulty out-of frame sequences from the library.

To perform the reading-frame selections we implemented a method published by Lutz and coworkers[13]. The procedure demands that bacteria are transformed with the library of interest. Unfortunately, due to experimental limitations in this transformation step, only about 10^7 individual variants can be sampled by the reading frame selection. The intended size of our full-length ADRB2 library, however, was 10^{12} to 10^{13} sequence variants (this upper limit of the sequence space that can be sampled is imposed by the ribosome display method). The reading-frame selection, therefore, had to be performed before the 12 building blocks were ligated to form the full-length library. To this end, the 12 building blocks were assembled to form six modules for reading-frame selection (colored in grey in **Fig. 2.5**) – none of them exceeded a theoretical library size of 10^7.

The reading frame selection, however, turned out to be very problematic. The control experiments failed to prove the concept. At this stage of the project, I teamed up with my coworker Hilmar Ebersbach to solve these problems. Ebersbach had worked on a similar project as mine. Within four months he had optimized the experimental parameters of the system and had established a robust reading-frame selection procedure.

In the meantime, however, I got engaged in experiments with my coworker Casim Sarkar. I found his studies on the directed evolution of GPCRs highly fascinating and he had made very good progress in the past couple of months. Since his time in our laboratory was coming to an end, it was agreed that I would continue his experiments. Eventually, these experiments kept me busy for the following years and they led to the studies presented in the **Chapters 4** and **5** of this thesis. My attention had gradually shifted away from the project on the water-soluble ADRB2 described in this chapter. At present, the project is temporarily stopped at the final stages of synthesizing the full-length library.

2.4 Materials and methods

Sequence analysis and homology modeling. Multiple sequence alignments (MSAs) were calculated with ClustalW using standard parameters. The MSAs were manually adjusted if necessary. Homology modeling was performed with Insight II software on a SGI Octane Workstation. The homology model of the human β_2 adrenergic receptor was build based on the crystal structure of bovine rhodopsin (PDB code 1f88). The model was restricted to the seven TM helices. The amino acid sequences of the seven TM helices of ADRB2 were aligned to the homologous regions of rhodopsin. The initial model was built by assigning the backbone coordinates of rhodopsin to the homologous positions of ADRB2. Then the side chains were assigned. The orientations of the side chains were optimized to remove steric clashes and to saturate hydrogen-bonding contacts by sampling the backbone-dependent rotamer library implemented in Insight II. Major steric clashes around TM2 and in the interface between TM3 and TM4 were removed by minor adjustments in TM2 and TM4. The solvent accessibly surface area for each amino acid in the model was calculated by rolling a probe (radius = 1.4 Å) over the receptor surface.

AGADIR. The AGADIR algorithm was run on peptides representing the designed TM helices. The following parameters were used: N-terminal = acetylation, C-terminal = amidation, temperature = 298 K, pH = 7.4, ionic strength = 0.2 M NaCl.

Synthesis of the gene library encoding water-soluble ADRB2 analogs. Unless stated otherwise, all experiments were performed as described[14]. DNA amplified by PCR was always purified by agarose gel electrophoresis. Oligonucleotides encoding the highly randomized positions in the gene library were synthesized by Metabion (Germany) by incorporating mixed trinucleotides[15] as building blocks (Glen Research, USA). The following mixtures of trinucleotides were used to encode a subset of the amino acids EKRQAVLF.

Table 2.1: Composition of the five trinucleotide mixtures encoding a subset of the amino acids EKRQAVLF (% values).

	Mix 1	**Mix 2**	**Mix 3**	**Mix 5**	**Mix 6**
E	34	34	34	32	10
K	17	17	17	16	30
R	17	17	17	16	30
Q	14	9	9	8	10
A	18	13	13	12	20
F				16	
L		10			
V			10		

All other oligonucleotides were from Microsynth (Switzerland). All PCRs were done with Phusion polymerase (Finnzymes). The restriction endonucleases BpiI and BsaI were from NEB (USA) and Fermentas (Canada).

The DNA of the highly randomized building blocks TM1, TM4, TM5, TM6 and TM7 (green bars in **Fig. 2.5**) was synthesized by four oligonucleotides; their sequences are shown below. The randomized positions in the oligonucleotide sequence are indicated by boldface font and the numbers (**1**, **2**, **3**, **5** and **6**) indicate the trinucleotide mixtures incorporated at a given position (for their amino acid composition, see **Table 1**). If degenerated genetic code – in addition to the trinucleotide mixtures – was used for randomization, it is indicated by the standard symbols also highlighted in boldface. For some fragments it was necessary to encode the randomized positions on more than one oligonucleotide (e.g. tm4_a). For clarity, the nucleotide segment encoding a given helix fragment is separated from the constant flanking segments by a space character. The experimental procedure for synthesizing an entire building block encoding a randomized helix is described in **Figure 2.6**.

Oligonucleotides for building block TM1:

```
tm1_a  5′-CGCAGCAGGAAGACGG GGTGGHAGSCATGGCG1CTGATG CGAGACCAGCAGAACGG-3′
tm1_b  5′-CGCAGCAGGAAGACGG GATG11ATTGTG2GCGA CGAGACCAGCAGAACGG-3′
tm1_c  5′-CGCAGCAGGAAGACGG GCGATT31GGCAAT CGAGACCAGCAGAACGG-3′
tm1_d  5′-CGCAGCAGGAAGACGG CAAT1CTGGTGATTAMGGCGA CGAGACCAGCAGAACGG-3′
```

Oligonucleotides for building block TM4:

```
tm4_a1 5′-CGCAGCAGGAAGACGG GACCRAAAACRAAGCGC CGAGACCAGCAGAACGG-3′
tm4_a2 5′-CGCAGCAGGAAGACGG GACCRAAAACCTGGCGC CGAGACCAGCAGAACGG-3′
tm4_a3 5′-CGCAGCAGGAAGACGG GACCRAACAGRAAGCGC CGAGACCAGCAGAACGG-3′
tm4_a4 5′-CGCAGCAGGAAGACGG GACCRAACAGCTGGCGC CGAGACCAGCAGAACGG-3′
tm4_b  5′-CGCAGCAGGAAGACGG GCGCRGGMAGTGATT66GTGT CGAGACCAGCAGAACGG-3′
tm4_c  5′-CGCAGCAGGAAGACGG GTGTGG1GTGKCGGCG CGAGACCAGCAGAACGG-3′
tm4_d  5′-CGCAGCAGGAAGACGG GGCG1ACCAGC5CTGC CGAGACCAGCAGAACGG-3′
```

Oligonucleotides for building block TM5:

```
tm5_a  5′-CGCAGCAGGAAGACGG CCAGGMATACGCG3GCGTCGGCG CGAGACCAGCAGAACGG-3′
tm5_b  5′-CGCAGCAGGAAGACGG GGCG13GCGTTTTAT2CCGC CGAGACCAGCAGAACGG-3′
tm5_c  5′-CGCAGCAGGAAGACGG CCGCTG13ATG31GTGT CGAGACCAGCAGAACGG-3′
tm5_d1 5′-CGCAGCAGGAAGACGG GTGTATGCGCGCGTGTTTCAAG CGAGACCAGCAGAACGG-3′
tm5_d2 5′-CGCAGCAGGAAGACGG GTGTATGCGCGCGTGCAGCAAG CGAGACCAGCAGAACGG-3′
```

Oligonucleotides for building block TM6:

```
tm6_aT 5′-CGCAGCAGGAAGACGG GCTGGMAACCCTGGCG3ATTATG CGAGACCAGCAGAACGG-3′
tm6_aV 5′-CGCAGCAGGAAGACGG GCTGGMAGTGCTGGCG3ATTATG CGAGACCAGCAGAACGG-3′
tm6_b  5′-CGCAGCAGGAAGACGG TATG1AYCTTTAMG2TGCTGG CGAGACCAGCAGAACGG-3′
tm6_c  5′-CGCAGCAGGAAGACGG CTGGCKTCCGTTCTTT1GTGAAC CGAGACCAGCAGAACGG-3′
tm6_d  5′-CGCAGCAGGAAGACGG GAAC33CATGTGAWAMAAGATA CGAGACCAGCAGAACGG-3′
```

Oligonucleotides for building block TM7:

```
tm7_a  5′-CGCAGCAGGAAGACGG GATAACCTGATTCSCAAAGAA CGAGACCAGCAGAACGG-3′
tm7_b  5′-CGCAGCAGGAAGACGG AGAAGHATATATT2CTGAACT CGAGACCAGCAGAACGG-3′
tm7_c  5′-CGCAGCAGGAAGACGG AACTGG1GGCTAT1AACA CGAGACCAGCAGAACGG-3′
tm7_d1 5′-CGCAGCAGGAAGACGG AACAGCGCGTTTAACCCGCWGATTT CGAGACCAGCAGAACGG-3′
tm7_d2 5′-CGCAGCAGGAAGACGG AACAGCGCGTTTAACCCGGMGATTT CGAGACCAGCAGAACGG-3′
```

The DNA for the two building blocks for TM2 and TM3 (brown bars in **Fig. 2.5**) were synthesized by assembly PCR using the following oligonucleotides:

Oligonucleotides for building block TM2:

```
tm2_1  5'-CGCAGCAGGAAGACGG TCTG1ATGGGC2GCCGTTGTGCCATTCGGTG-3'
tm2_3a 5'-ATGAAGATGTGGACGTTCGGCAACSAGTGGTGTGAG5TGGA CGAGACCAGCAGAACGG-3'
tm2_3b 5'-ATGAAGATGTGGACGTTCGGCAACTTCTGGTGTGAG5TGGA CGAGACCAGCAGAACGG-3'
```

Oligonucleotides for building block TM3:

```
tm3_1  5'-CGCAGCAGGAAGACGG TGGACCKCGATCGACGTCTTATGCGTCACGGCCAGCATTGAAACCC-3'
tm3_2  5'-CGGGCTCGTAATCGCCWGATAGCGGTCTDCCGCAATGACACCACGGGTTTCAATGCTGGCC-3'
tm3_3  5'-GGCGATTACGAGCCCGTTCAAATACCAAAGCTTACTGACC CGAGACCAGCAGAACGG-3'
```

The following primers were used to generate double-stranded DNA from the above oligonucleotides by PCR. They anneal to the constant part of each randomized oligonucleotide.

```
Bpi_for 5'-CGCAGCAGGAAGACGG-3'
Bsa_rev 5'-CCGTTCTGCTGGTCTCG-3'
```

The double-stranded DNA for the building blocks N-ter, IL1, EL2, IL3 and C-ter (blue bars in **Fig. 2.5**) was synthesized by Geneart (Germany) and provided in a plasmid format. These building blocks were amplified by PCR from the plasmid and ligated to other building blocks according to the scheme in **Figure 2.5**.

2.5 Conclusions and future directions

The conversion of the β_2 adrenergic receptor into a water-soluble receptor analog presents a very challenging protein-engineering problem. In this chapter, I have described a strategy towards solving this problem. It is based on the principles of directed evolution, namely mutagenesis and selection. This chapter describes the redesign of the phospholipid-contacting surface of ADRB2 and the implementation of the redesign by the synthesis of a combinatorial library on the DNA level. Although the final proof that a GPCR can be converted into a water-soluble analog has not yet been demonstrated here, the results presented in this study provide a strong experimental framework towards this goal.

The future directions in this project comprise two main steps. First, in order to complete the synthesis of the combinatorial library, the six reading-frame selection modules have to undergo reading-frame selection based on the experimental conditions proposed by Hilmar Ebersbach. The six modules can then be assembled to form a high-quality full-length library. Second, the library should be selected for ligand binding competence by ribosome display in order to isolate ADRB2 receptor analogs that adopt the 7TM fold in aqueous solution.

2.6 References

1. Frank, S., Kammerer, R.A., Hellstern, S., Pegoraro, S., Stetefeld, J., Lustig, A., Moroder, L. & Engel, J. **Toward a high-resolution structure of phospholamban: design of soluble transmembrane domain mutants**. *Biochemistry* **39**, 6825-6831 (2000).
2. Li, H., Cocco, M.J., Steitz, T.A. & Engelman, D.M. **Conversion of phospholamban into a soluble pentameric helical bundle**. *Biochemistry* **40**, 6636-6645 (2001).
3. Slovic, A.M., Summa, C.M., Lear, J.D. & DeGrado, W.F. **Computational design of a water-soluble analog of phospholamban**. *Protein Sci* **12**, 337-348 (2003).
4. Slovic, A.M., Kono, H., Lear, J.D., Saven, J.G. & DeGrado, W.F. **Computational design of water-soluble analogues of the potassium channel KcsA**. *Proc Natl Acad Sci U S A* **101**, 1828-1833 (2004).
5. Slovic, A.M., Stayrook, S.E., North, B. & Degrado, W.F. **X-ray structure of a water-soluble analog of the membrane protein phospholamban: sequence determinants defining the topology of tetrameric and pentameric coiled coils**. *J Mol Biol* **348**, 777-787 (2005).
6. Mitra, K., Steitz, T.A. & Engelman, D.M. **Rational design of 'water-soluble' bacteriorhodopsin variants**. *Protein Eng* **15**, 485-492 (2002).
7. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M. & Miyano, M. **Crystal structure of rhodopsin: A G protein-coupled receptor**. *Science* **289**, 739-745 (2000).
8. Cherezov, V., Rosenbaum, D.M., Hanson, M.A., Rasmussen, S.G., Thian, F.S., Kobilka, T.S., Choi, H.J., Kuhn, P., Weis, W.I., Kobilka, B.K. & Stevens, R.C. **High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor**. *Science* **318**, 1258-1265 (2007).
9. Ballesteros, J.A., Shi, L. & Javitch, J.A. **Structural mimicry in G protein-coupled receptors: implications of the high-resolution structure of rhodopsin for structure-function analysis of rhodopsin-like receptors**. *Mol Pharmacol* **60**, 1-19 (2001).
10. Rasmussen, S.G., Choi, H.J., Rosenbaum, D.M., Kobilka, T.S., Thian, F.S., Edwards, P.C., Burghammer, M., Ratnala, V.R., Sanishvili, R., Fischetti, R.F., Schertler, G.F., Weis, W.I. & Kobilka, B.K. **Crystal structure of the human beta2 adrenergic G-protein-coupled receptor**. *Nature* **450**, 383-387 (2007).
11. Munoz, V. & Serrano, L. **Elucidating the folding problem of helical peptides using empirical parameters**. *Nat Struct Biol* **1**, 399-409 (1994).
12. Lacroix, E., Viguera, A.R. & Serrano, L. **Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters**. *J Mol Biol* **284**, 173-191 (1998).
13. Gerth, M.L., Patrick, W.M. & Lutz, S. **A second-generation system for unbiased reading frame selection**. *Protein Eng Des Sel* **17**, 595-602 (2004).
14. Sambrook, J., Fritsch, E.F. & Maniatis, T. **Molecular Cloning: a Laboratory Manual, 2nd edit.** *Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY* (1989).

15. Virnekäs, B., Ge, L., Plückthun, A., Schneider, K.C., Wellnhofer, G. & Moroney, S.E. **Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis**. *Nucleic Acids Res* **22**, 5600-5607 (1994).

Chapter 3

Engineering and functional immobilization of opioid receptors

3.1 Published article

Protein Engineering, Design & Selection vol. 18 no. 3 pp. 153–160, 2005
Published online March 24, 2005 doi:10.1093/protein/gzi012

Engineering and functional immobilization of opioid receptors

David Ott[1], Yvonne Neldner, Régis Cèbe[2], Igor Dodevski and Andreas Plückthun[3]

Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

[1]Present address: Universitätsspital Zürich, Departement Neuropathologie, Schmelzbergstr. 12, CH-8091 Zurich, Switzerland
[2]Present address: Novartis Pharma AG, WSJ 88/7.07, CH-4000 Basel, Switzerland

[3]To whom correspondence should be addressed.
E-mail: plueckthun@bioc.unizh.ch

Opioid receptors, like many G protein-coupled receptors (GPCRs), are notoriously unstable in detergents. We have now developed a more stable variant of the μ-opioid receptor (MOR) and also a method for the immobilization of solubilized, functional opioid receptors on a solid phase (magnetic beads). Starting with the intrinsically more stable κ-opioid receptor (KOR), we optimized the conditions (i.e. detergents and stabilizing ligands) for receptor extraction from lipid bilayers of HEK293T cells to obtain maximal amounts of functional, immobilized receptor. After immobilization, the ligand binding profile remains the same as observed for the membrane-embedded receptor. For the immobilized wild-type μ-opioid receptor, however, no conditions were found under which ligand binding capacity was retained. To solve this problem, we engineered the receptor chimera KKM where the N-terminus and the first transmembrane helix (TM1) of wild-type MOR is exchanged for the homologous receptor parts of the wild-type KOR. This hybrid receptor behaves exactly as the wild-type MOR in functional assays. Interestingly, the modified MOR is expressed at six times higher levels than wild-type MOR and is similarly stable as wild-type KOR after immobilization. Hence the immobilized MOR, represented by the chimera KKM, is now also amenable for biophysical characterization. These results are encouraging for future stability engineering of GPCRs.
Keywords: functional immobilization/μ-opioid receptor/opioid receptors

Introduction

Opioid receptors are members of the peptide-binding subfamily of class A of G protein-coupled receptors (GPCRs). Agonists selective for the κ-opioid receptor (KOR) produce an effective analgesia without the substantial side effects (constipation, respiratory depression, vomiting and physical dependence) associated with μ-opioid receptor (MOR)-selective agonists such as morphine (Reece *et al*., 1994). MOR represents a key mediator for a wide range of both beneficial and adverse activities of morphine. Using knockout mice, MOR was shown to be necessary for a reward of and dependence on other drugs of abuse [e.g. Δ^9-tetrahydrocannabinol and ethanol; reviewed elsewhere (Gaveriaux-Ruff and Kieffer, 2002)]. The detailed mechanistic and, ideally, structural understanding of opioid receptors and their interactions with agonists and antagonists are therefore of obvious importance.

Detergent-mediated extraction of receptors from membranes and subsequent purification of the receptor in the presence of detergents are prerequisites for any direct high-resolution study of receptor structure by X-ray crystallography or NMR and even for detailed biochemical characterization of the purified protein (Clark *et al*., 2001; Karlsson and Löfas, 2002; Banères and Parello, 2003; Banères *et al*., 2003; Stenlund *et al*., 2003). Owing to the low stability of most GPCRs in detergent, one could up to now, however, investigate their structure only indirectly and at an unsatisfactory level of detail. The single exception, rhodopsin, may be atypical in some respects and was studied after isolation from its natural source, where it is highly abundant (Palczewski *et al*., 2000). Most antagonistic or agonistic ligands for therapeutic intervention have thus been obtained from whole-cell screening (see, e.g., Conway and Demarest, 2002; Kassack *et al*., 2002; Wise *et al*., 2004).

One possibility is to follow changes of receptor conformation, albeit at low spatial resolution, in real time as a function of ligand-induced receptor activation, e.g. by fluorescence microscopy (Neumann *et al*., 2002). The receptor is first solubilized from the plasma membrane of the expression host and then immobilized on a solid phase. Then a fluorescence label is coupled to a cysteine residue introduced into a desired position in the receptor sequence. However, not even this analysis can be universally applied, as some receptors lose their native structure too fast in detergents.

While detergent extraction and maintenance of activity were possible with KOR, preliminary experiments showed that the homologous MOR was significantly less stable upon solubilization with detergents. We were therefore interested in finding the molecular determinants for their different behavior. We report here the results for hybrid proteins, leading to an engineered MOR with substantially increased expression level and higher stability in detergents. We believe that this approach can be of general utility.

We also present a setup for the functional immobilization of GPCRs and demonstrate it for opioid receptors. Opioid receptors are known to be unstable in detergents [reviewed elsewhere (Simon, 1986, 1991; Smith and Loh, 1991; Li *et al*., 1993)] and we therefore sought an optimal balance between functionality (such as ligand binding) and harsher extraction conditions leading to homogeneity of the immobilized receptor. In order to characterize qualitatively the immobilized receptor, we compared the ligand binding profiles of immobilized and membrane-embedded receptor.

Materials and methods

Materials

Most detergents were obtained from Anatrace. *rac*-2,3-Dihydroxypropyloctyl sulfoxide (OPSO) was purchased from

Bachem, non-detergent sulfobetaines (NDSB series) and Zwittergent 3-14 from Calbiochem, sodium lauroylsarcosine (sarcosyl), sodium cholate, sodium dodecyl sulfate (SDS), Lubrol PX ($C_{12}E_9$) and Nonidet P-40 (NP-40) from Sigma and Triton X-100 and Tween-20 from Fluka. Cholesteryl hemisuccinate (Cat. No. C-6013; CHS; 1% stock solution prepared in 5% CHAPS), the opioid ligands butorphanol (Cat. No. B-9156), naltrexone (Cat. No. N3136), nalbuphine (Cat. No. N-4396) and a protease inhibitor cocktail [Cat. No. P-8340; used at 0.3% (v/v) final concentration] were supplied by Sigma and Dermorphin (Cat. No. H-2565) by Bachem. Sources of all other chemicals used in this study are indicated elsewhere (Ott *et al*., 2004).

Construction of receptor variants

Genes of all opioid receptor variants (wild-types and chimeras) were cloned in frame with an N-terminal FLAG-tag and C-terminal myc- and His-tags. The cloning of the wild-type KOR (Swiss-Prot No. P34975) and FCL KOR (Ott *et al*., 2004) genes into the vector pcDNA3.1a(–)/myc/His was described previously (Ott *et al*., 2004). The gene for the wild-type (wt) MOR (Swiss-Prot No. P33535) was inserted into the vector pcDNA3.1a(–)/myc/His with the N-terminal FLAG-tag (MDYKDDDDK), resulting in a short linker (Gln–Leu) between the FLAG-tag and the receptor. We designed three receptor chimeras for the MOR, where either the N-terminal extracellular part or the first transmembrane helix (TM1) or both parts together were exchanged for the homologous sequences of the KOR. The decisions on the domain boundaries were made based on the predictions in the SWISS-PROT Annotated Protein Sequence Database entry files for the receptors (see Results). For the chimera KKM, residues Pro4–Thr88 of the KOR were N-terminally fused to residues Lys98–Pro398 of the MOR; for the chimera KMM, residues Pro4–Ile58 of KOR were N-terminally fused to residues Met65–Pro398 of the MOR and for the MKM chimera, residues Pro59–Thr88 of KOR replaced residues Met65–Thr97 in MOR.

Cell culture, transient transfection, radioligand binding assays, membrane preparation, western blot analysis and [^{35}S]GTPγS binding assay

These techniques were performed essentially as described previously (Ott *et al*., 2004). We routinely centrifuged a stock solution of [^{35}S]GTPgS (50 nM in Tricine buffer, pH 7.4, with 10 mM DTT) for 1 h at 4°C at 90 000 r.p.m. (400 000 *g*) in a Beckman Optima TLX ultracentrifuge in order to remove any insoluble impurities which might interfere with the assay.

Immobilization of receptor on paramagnetic beads

Cell pellets of transfected HEK293T cells (from one well in a 24-well plate) were resuspended in 100 µl of immunoprecipitation (IP) buffer [TBS with a CHAPS–cholesteryl hemisuccinate (CHS) mixture (8.3/1.7 mM), 10% glycerol and 10 mM $MgCl_2$] and incubated for 10 min on ice. The suspension was centrifuged for 10 min at 3300 r.p.m. (1100 *g*) in an Eppendorf 5417R centrifuge at 4°C. The supernatant was mixed with the anti-myc monoclonal antibody 9B11 [Cell Signaling, Cat. No. 2276; diluted 1:200 (v/v)] and paramagnetic Dynabeads carrying protein G (Dynal, Cat. No. 100.03; 10 µl/sample). The suspension was shaken (1400 r.p.m.) for 30 min at 30°C or for 1 h at 4°C. By using a magnetic rack, beads were washed four times with 300 µl of IP buffer at room temperature (RT) (with 3 min incubations between the washes). For western blot analyses, beads were resuspended in 30 µl of 2.5-fold concentrated SDS–PAGE loading buffer and incubated for 10 min at 42°C. For the binding assay, beads were resuspended in 200 µl of TBS containing CHAPS–CHS (1.7/0.33 mM) and incubated with radioligand in the presence or absence of non-labeled competitor for 1 h at RT. Separation of free and bound ligand was performed on a filtration vacuum manifold (Millipore) as described elsewhere (Ott *et al*., 2004) with the following modifications: after the filtration, beads on glass-fiber filters were transferred into 20 ml scintillation vials (Ratiolab), scintillation cocktail (5 ml) was added and 16 h later the samples were measured in a BETAmatic β-counter (from Kontron).

When various detergents were tested, the complete detergent-containing IP buffer including protease inhibitor cocktail was also used for the binding assay. Extraction of receptor from cells by using sarcosyl (0.5%), SDS (0.15%), sodium deoxycholate (0.5%) and hexyl maltoside (2%) resulted in viscous cell lysates. Addition of DNase I (Roche) and a short incubation for 15 min at 30°C partially helped to reduce the viscosity of the lysate. The viscosity could not be reduced in sodium deoxycholate (0.5%) lysate and therefore we excluded this detergent from screening. Sarcosyl (0.5%) and particularly CHAPSO (0.6%) were prone to precipitation at 4°C. Therefore, we used the IP buffers containing these detergents at RT for the washing steps in the radioligand binding assays. When the receptor was immunoprecipitated in digitonin (1%), IP buffer with 1 M NaCl was used.

The relative stability of receptor variants in detergents was determined as the ratio of specific ligand binding measured on beads with the immobilized receptor and on whole cells. The relative receptor stabilities were determined only for conditions where receptor variants were expressed in the absence of ligand in the growth medium.

Results

Engineering a stabilized μ opioid receptor

We wished to establish a general strategy to solubilize GPCRs in detergent and immobilize them for functional studies. In preliminary experiments we observed that the rat KOR lost its activity in detergent at a much slower rate than the highly homologous MOR. At the sequence level, MOR and KOR show 57% pair-wise sequence identity with the highest divergence at the N-terminus (20% identical amino acid positions). We wished to exploit these observations to find a strategy for stabilizing MOR.

There are several reports that the N-terminus, including the first transmembrane helix (TM1), of opioid receptors is not important for receptor specificity towards subtype-selective ligands (Onogi *et al*., 1995; Wang *et al*., 1995; Ide *et al*., 2000). To test whether the N-terminus is responsible for the higher stability of immobilized KOR in detergent, we engineered a receptor chimera KKM, where the N-terminus and TM1 up to the first cytoloop originates from wt KOR and the rest of the sequence is taken from wt MOR (Figure 1).

We first determined the ligand binding profiles of the constructs MOR, KOR and KKM in membrane preparations of HEK293T cells expressing the receptor variants.

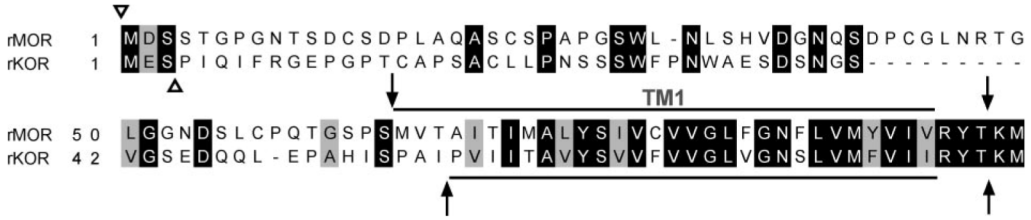


Fig. 1. Sequence alignment of the relevant region in the μ- and κ-opioid receptors from rat. The aligned residues of the N-terminal extracellular domain and transmembrane helix 1 (TM1) up to the first cytoplasmic loop are shown. The overscored and underscored regions represent the predicted TM1 according to the SWISS-PROT Annotated Protein Sequence Database entries for the receptors. For the construction of the chimeric receptors, we used the cutting sites indicated by arrows. The N-terminal FLAG-tag was introduced at the sites marked by an open triangle. The sequences were aligned using Clustal W (version 1.81) followed by manual refinement. SWISS-PROT accession numbers for the sequences are P33535 for rMOR and P34975 for rKOR.

Table I. Ligand binding profiles of membrane-embedded (M) or immobilized (I) receptor variants[a]

Ligand	Receptor variant							
	wt MOR		KKM		wt KOR		FCL KOR	
	M	I	M	I[e]	M	I[f]	M	I[g]
Diprenorphine	0.3 ± 0.2	ND[c]	0.4 ± 0.2	0.07 ± 0.04	0.4 ± 0.1[b]	0.12 ± 0.02	0.6 ± 0.1[b]	0.5 ± 0.3
U-50488	>1000	ND	>1000	>1000	7 ± 7	5.3 ± 0.4	70[d]	70 ± 60
DAMGO	30 ± 10	ND	58 ± 1	70 ± 50	>1000[b]	>1000	>1000[b]	>1000
Dermorphin	40 ± 30	ND	60 ± 30	24 ± 6	>1000	>1000	ND	ND
Naloxone	3 ± 1	ND	2 ± 1	0.9 ± 0.4	4 ± 2[b]	4 ± 1	6 ± 2[b]	10 ± 7
Butorphanol	0.5 ± 0.1	ND	0.6 ± 0.2	1.9 ± 0.4	0.23 ± 0.03	0.7 ± 0.4	9[d]	8 ± 11

[a]Equilibrium dissociation constant for [^{3}H]diprenorphine and equilibrium inhibition constants for all other tested ligands are expressed in nM and were determined as described previously (Ott *et al*., 2004).
[b]Data determined previously (Ott *et al*., 2004).
[c]ND, not determined.
[d]Value obtained from one independent experiment.
[e]Receptor was expressed in the presence of naltrexone (20 μM) in the growth medium.
[f]Receptor was expressed in the presence of naloxone (20 μM) in the growth medium.
[g]Receptor was expressed in the presence of butorphanol (20 μM) in the growth medium.

The equilibrium binding constants are summarized in Table I. Notably, the wt MOR and the chimera KKM show similar pharmacological profiles for the ligands tested (opioid antagonists [^{3}H]diprenorphine and naloxone, the non-selective agonist butorphanol, the κ-selective agonist U-50488 and μ-selective agonists DAMGO and dermorphin). Since all three constructs showed comparable affinities for the radioligand [^{3}H]diprenorphine, we decided to perform all further competitive binding assays at non-saturating concentrations of the radioligand (slightly above K_D), where we measured high signal-to-noise ratios. We could thus avoid working at saturating concentrations, which would give rise to high non-specific binding signals. In this experimental setup, the presence of functional receptor (either on whole cells or in solubilized, immobilized form) is directly determined via its capacity to bind radioligand.

We compared relative expression levels of wt KOR, wt MOR and the KKM chimera by using whole cell binding assays with the radioligand [^{3}H]diprenorphine (1 nM). MOR was functionally expressed at slightly lower levels, i.e. ∼50% of the KOR expression. Interestingly, the chimera KKM is expressed at a ∼3-fold higher level than KOR and hence at ∼6-fold higher levels than MOR (Figure 2).

In order to assess a coupling of the receptor variants to G-proteins, we used the [^{35}S]GTPγS binding assay with the membrane preparations. The potency of the μ-selective agonist dermorphin to activate G-proteins via the receptor is very similar for the wt MOR and the chimera KKM (half-maximal activity ∼20 nM; Figure 3). Taking these data together, the chimera KKM behaves like the wt MOR in all functional assays performed.

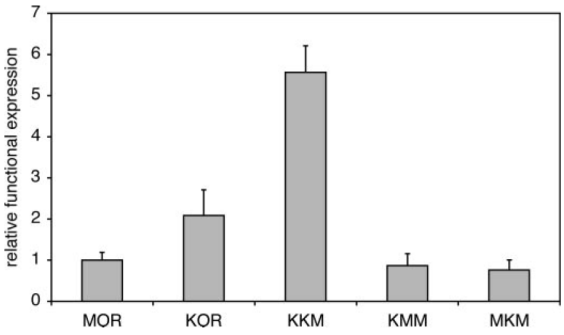


Fig. 2. Functional expression of receptor variants in HEK293T cells. Specific binding of receptor variants is given in relative numbers, with the wt MOR arbitrarily set to 1. Transfected cells were grown in the absence of receptor ligand in the medium. The binding of antagonist [^{3}H]diprenorphine (1 nM) was measured on whole cells. This experiment represents the average of five independent receptor expression experiments for MOR, KOR and KKM and three independent receptor expression experiments for KMM and MKM (duplicate or triplicate samples were analyzed in each independent experiment).

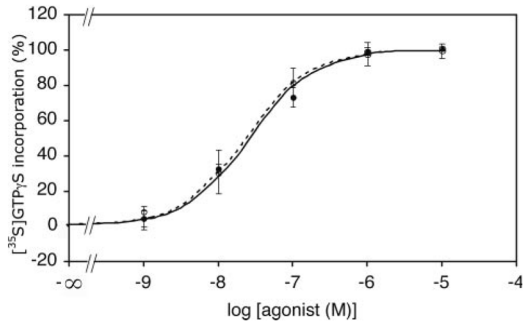


Fig. 3. G-protein coupling for the wt MOR and chimera KKM measured by [^{35}S]GTPγS binding assays. Membranes were prepared from cells expressing receptor in the presence of naloxone and [^{35}S]GTPγS binding assays were performed in the presence of various concentrations of the agonist dermorphin as described in Materials and methods. Experimental data were normalized by taking the basal activation in the absence of ligand as a minimum (0%) and the activation caused by 10 μM dermorphin as a maximum (100%). Data are represented for the wt MOR, by filled circles and the solid fitted curve and for the chimera KKM by open circles and the dashed fitted curve.

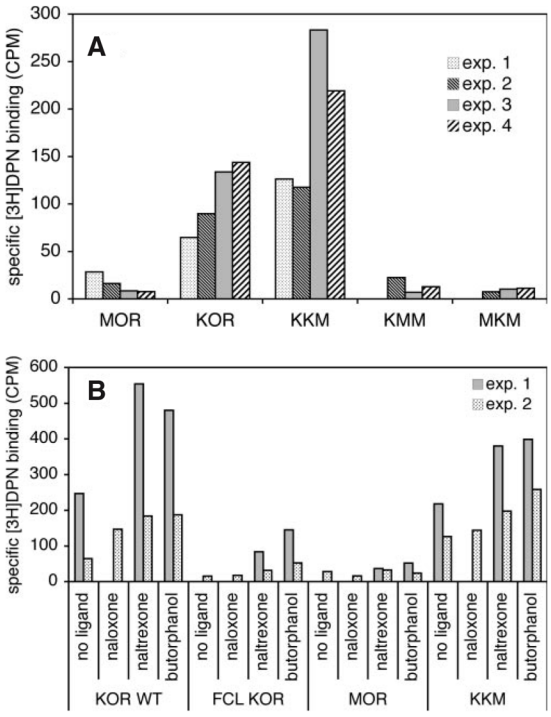


Fig. 4. Functional immobilization of opioid receptors. Receptor, expressed in the absence or presence of various ligands in the growth medium, was extracted from HEK293T cells in CHAPS–CHS, immobilized on paramagnetic beads and tested for specific [^{3}H]diprenorphine binding as described in Materials and methods. Bars of different texture represent independent experiments. (**A**) Receptor variants were expressed in the absence of ligand in the growth medium. (**B**) Receptor variants were expressed in the absence or presence of various ligands in the growth medium. The effect of ligands in the growth medium on the functionality of the immobilized receptor was tested in two independent experiments (each with duplicate samples), with the exception of naloxone, which was tested in one independent experiment with duplicate samples only.

We then investigated the solubilization and immobilization of KOR, MOR and KKM with the strategy described in the Materials and methods section. We found that by using the optimized solubilization and immobilization format, both KOR and MOR can be efficiently immobilized on the beads, as detected by western blots (not shown). KOR was immobilized in an active form, whereas the immobilized MOR was almost non-functional (Figure 4A). In sharp contrast, the KKM chimera was immobilized in an active form at even higher levels than wt KOR (Figure 4A).

The substantial difference in the yield of immobilized receptor between the engineered receptor form KKM and MOR can be attributed to two factors, expression level and amount of functional protein in detergent after solubilization: KKM was expressed in cells at a substantially higher level (~6-fold) than MOR (Figure 2) and the relative stability of KKM (7 ± 2%, n = 4), i.e. the percentage of receptor molecules remaining active after immobilization, was also higher than the relative stability of MOR (4 ± 3%, n = 4). For comparison, the relative stability of KOR was 11 ± 2% (n = 4). These data indicate that it was indeed possible to improve a wild-type GPCR by engineering for both higher cellular expression and stability in detergent.

We then wanted to clarify whether the favorable properties of KKM arise from the N-terminal region and/or TM1 of the receptor. For this purpose, we separately exchanged the N-terminus and TM1 between KOR and MOR. We created the constructs KMM and MKM, where the first, second and third letters denotes the origins of the N-terminal domain, TM1 and the remainder of the protein, respectively (Figure 1).

The equilibrium binding constants for radioligand binding to KMM and MKM (measured with membrane preparations) was comparable to wt MOR ($K_D \approx 0.3$ nM; data not shown), consistent with the expectation that the major part of the receptor determines the binding site. The expression of KMM and MKM was determined in parallel with the receptors MOR, KOR and KKM using binding assays on whole cells. We found that the expression levels of KMM and MKM were as low as for wt MOR and that these chimeras could not equal the expression levels of the KKM construct (Figure 2).

Furthermore, the immobilized chimeras KMM and MKM were almost non-functional, similarly to the wt MOR (Figure 4A). Hence we cannot narrow down the favorable effect seen in KKM to particular residues or receptor domains. However, it appears that the simultaneous exchange of the N-terminal domain and the TM1 resulted in a very favorable situation and only the KKM chimera of the MOR could be functionally immobilized with an appreciable yield.

GPCR solubilization and immobilization

We developed a strategy of GPCR solubilization and immobilization as a basis for functional and mechanistic receptor characterization *in vitro*. In order to characterize the functionality of the solubilized receptor, we found it important to compare the ligand binding affinities of receptor variants determined in the presence of detergent with those determined for the receptor embedded in the native membranes (whole cell binding assay). If both could be done on a solid phase, binding assays would become greatly simplified. Our goal was to immobilize the solubilized opioid receptors in highly enriched

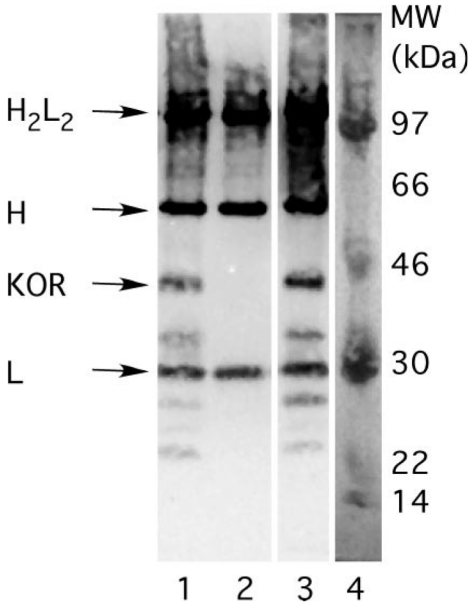


Fig. 5. Immunoprecipitation of KOR variants. KOR fused to FLAG-, myc- and His-tags was extracted from transfected HEK293T cells and immobilized on paramagnetic beads by using the anti-myc antibody in the IP buffer containing CHAPS (10 mM). Residual proteins on the beads were then analyzed by SDS–PAGE and western blots by using the anti-tetra-His antibody as described in Materials and methods. Results obtained with the cells transfected with wt KOR (lane 1), no DNA (lane 2) or KOR FCL (lane 3) are shown. Lane 4 corresponds to Rainbow marker RPN756 (Amersham). Several bands on the blot correspond to various forms of the anti-myc antibody (depicted by arrows), such as reduced and dissociated light (L) or heavy (H) chains or whole IgG (H_2L_2). The anti-myc antibody used for immobilization is of murine origin and is therefore detected on the blot by the secondary (anti-murine) antibody used for detection of the primary murine anti-tetra His antibody. The band, which corresponds to the full-length non-glycosylated form of KOR (~43 kDa), is also indicated by an arrow. All other protein bands in lanes 1 or 3 that are not observed in lane 2 correspond to degradation products of the receptor.

form in an oriented and functional fashion. We carried out immobilization studies at a temperature of 30°C, which is suitable also for fluorescence and signaling studies.

Preliminary experiments had indicated that the anti-myc tag antibody could efficiently immunoprecipitate KOR, carrying FLAG- and myc-tags, in buffer containing the detergent CHAPS (Figure 5), and also under relatively harsh conditions such as the radioimmunoprecipitation assay (RIPA) buffer (not shown). Therefore, we decided to use the anti-myc antibody for further optimization of the IP procedure. It was necessary to test a variety of detergents for maximal receptor extraction and maintenance of activity and to investigate the influence of stabilizing ligands during the solubilization and immobilization steps.

Choice of detergent. First, we expressed wt KOR in HEK293T cells in order to test the suitability of a variety of detergents for the solubilization and immobilization of the receptor in functional form. CHAPS was previously reported as a suitable detergent for the solubilization of opioid receptors when solubilization was followed by receptor reconstitution in liposomes (Simon, 1986; Zukin and Maneckjee, 1986; Smith and Loh, 1991; Gioannini *et al*., 1993; Fan *et al*., 1995). When screening for a suitable detergent, we found that 10 mM CHAPS was the sole condition under which wt KOR was immobilized in a highly enriched and functional form. While the use of other detergents for the IP also resulted in receptor enrichment (Figure 6), specific ligand binding was not detected or was significantly lower than for the conditions with CHAPS (data not shown). These inactivating detergents tested included sarcosyl (0.5%), sodium cholate (0.5%), Cymal-5 and Cymal-6 (both 0.5%), octyl glucoside (1%), Triton X-100 (0.05%), Nonidet P-40 (1%), LDAO (0.1%), octyl and nonyl maltoside (both 1%), sucrose monodecanoate (0.5%), $C_{12}E_9$ (0.2%), HEGA-10 (0.5%), digitonin (1%), Fos-choline-14 (0.14%), Fos-choline-16 (0.2%), dodecyl maltoside (0.3%) and Zwittergent 3-14 (0.2%).

Table II. Ligand binding detected for KOR wt immobilized in various reagents

Detergent	Concentration of detergent (%)	Relative specific binding (%)
CHAPS	0.6	100
None	–	200 ± 60
NDSB-195	0.5	300 ± 60
NDSB-195	2	170[a]
NDSB-201	0.5	270 ± 60
NDSB-211	0.5	320 ± 40
NDSB-211	2	180[a]

[a]Value obtained from one independent experiment.

When we performed receptor IPs in the absence of detergent or in the presence of hexyl maltoside (2%) or non-detergent sulfobetaines (NDSB-195, NDSB-201 or NDSB-211; all at 0.5%), the measured specific ligand binding was higher than what we observed for IP in CHAPS (Table II). However, receptor bands (representing the full-length and degradation products of KOR) were very faint on western blots (Figure 6A, lane 5) and preparations of the immobilized receptor were contaminated with other cellular proteins (Figure 6B, lane 5). These results indicate that membrane fragments were deposited on the beads and that these reagents were not able to solubilize the receptor sufficiently from membranes.

When we used OPSO (1%), dodecyl maltoside (0.03%), Fos-choline-14 (0.014%) or Zwittergent 3-14 (0.02%) at concentrations only 2–3 times higher than their theoretical critical micelle concentration (CMC), receptor bands (representing the full-length and degradation products of KOR) were also very faint on western blots (results were comparable to Figure 6A, lane 5) and preparations of the immobilized receptor were again contaminated with many cellular proteins (results were comparable to Figure 6B, lane 5). Furthermore, no functional receptor was detected in these preparations. Conversely, some detergents result in a prominent receptor band (shown for LDAO as an example, Figure 6B, lane 6), but no activity way found (see enumeration of inactivating detergents above). These results indicate that most detergents were deleterious for receptor function, some even at low concentrations which were not even sufficient for the complete release of receptor from membranes into detergent micelles.

CHS in a mixture with CHAPS and dodecyl maltoside was successfully used for the solubilization of the neurotensin receptor (Tucker and Grisshammer, 1996). Because dodecyl maltoside is deleterious to KOR (see above), we tested a mixture of CHAPS (0.5%) and CHS (0.1%) for the IP of wt KOR.

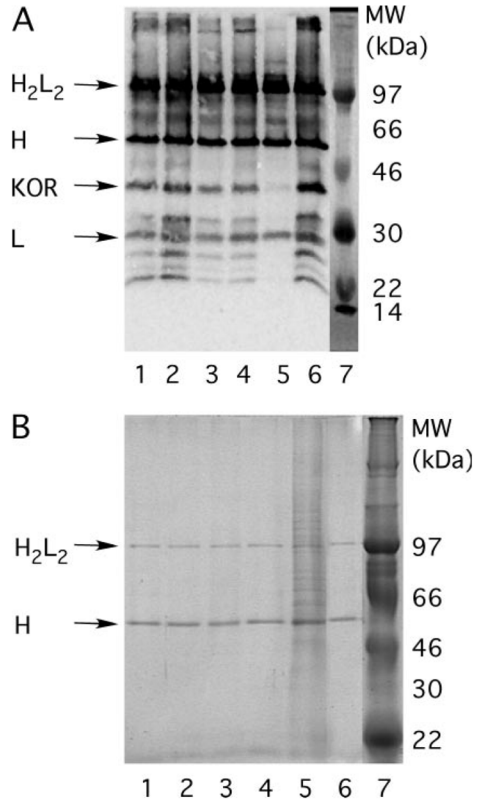


Fig. 6. Immunoprecipitation of wt KOR in various detergents. KOR extracted from transfected HEK293T cells was immobilized on paramagnetic beads by using the anti-myc antibody. Residual proteins on the beads were then analyzed by SDS–PAGE and western blot by using the anti-tetra-His antibody as described in Materials and methods. Various detergents were used for IP of the receptor and only a subset is shown. For a full account, see the text. (**A**) Western blot; (**B**) SDS–PAGE gel stained with Coomassie Brilliant Blue. Lane 1, CHAPS (0.6%); lane 2, CYMAL-5 (0.5%); lane 3, octyl glucoside (1%); lane 4, Nonidet P-40 (1%); lane 5, no detergent; lane 6, LDAO (0.1%); lanes A7 and B7, Rainbow marker RPN756 (Amersham).

Interestingly, the addition of CHS to the CHAPS-containing IP buffer resulted in a 3-fold increase in the yield of functional receptor. The purity of the receptor immobilized in CHAPS–CHS or only CHAPS did not differ (data not shown). We therefore selected the detergent mixture CHAPS–CHS for the setup of the IP procedure also for other opioid receptor variants. By using CHAPS–CHS, we achieved the best compromise between wt KOR homogeneity and functionality.

As described above, MOR was immobilized in CHAPS–CHS only in non-functional form. The same observation was made for the detergents Cymal-5 (0.5%) and dodecyl maltoside (0.3%), where no specific ligand binding to MOR was detected. In contrast, IP of the chimera KKM led to the immobilization of receptor in functional form, and in CHAPS–CHS about 3-fold more functional receptor was obtained than by IP in CHAPS alone. The chimera KKM immobilized in dodecyl maltoside (0.3%) or Cymal-5 (0.5%) was not functional (data not shown).

Choice of ligand. The addition of ligand to the receptor before solubilization can have two effects. First, when it is added during cell growth, it may lead to an improved receptor expression, possibly by preventing receptor internalization. Such an effect was seen for the FCL (free-cysteine-less; a receptor devoid of all free cysteines) mutant of KOR (Ott *et al.*, 2004) and also for other receptors (Li *et al.*, 2001; McLean *et al.*, 2002). Second, it may stabilize the receptor during the solubilization step, when the ligand is added to the cells before detergent addition. Favorable effects were observed for MOR (Weems *et al.*, 1996) and endothelin B receptor (Doi *et al.*, 1997). Our intention was to keep a maximal amount of receptor occupied with ligand in order to stabilize the receptor during solubilization. For this reason, we routinely used mild washing conditions for the transfected cells during the cell harvest (three washes with PBS at 4°C with ~2 min incubations between the washes). Provided that the ligand has a high (subnanomolar) affinity to its receptor and hence a very slow dissociation rate, it may block to some extent the access of radioligand to the receptor during the following binding assay. Very long incubations would be necessary to remove the bound ligand from the binding site on the receptor. Binding assays performed on whole cells grown in the presence of ligand may therefore underestimate the true receptor expression. Therefore, it is very difficult to make statements about the relative stability of receptor variants that have been exposed to ligands before solubilization. Nevertheless, one can measure the beneficial effect of various ligands on increasing the yield of functional immobilized receptor, being aware that these measurements may underestimate the true contributions of ligands on expression, as the effects are partially masked.

We tested the effect of naloxone and other ligands, present in the growth medium (at a concentration of 10 μM) during receptor expression, on the yield of the functional receptor immobilized in the detergent mixture CHAPS–CHS. We observed that the wt KOR and the chimera KKM can be immobilized in functional form in the CHAPS–CHS mixture, even when expressed in the absence of ligand. Nevertheless, the presence of various ligands during receptor expression leads to an increase in the observed amount of active, solubilized receptor. The ligands naloxone, naltrexone and butorphanol are most beneficial for increasing the amount of functional immobilization of wt KOR. Similarly, the ligands naltrexone, U-50488 and butorphanol improved most significantly the amounts of the functionally immobilized chimera KKM (Figure 4B). The highest amounts of functionally immobilized FCL KOR were observed when the ligands naltrexone or butorphanol were present in the growth medium. Nevertheless, it is obvious that the yield of the functional immobilized FCL KOR is significantly lower than for the wt KOR (Figure 4B). In contrast, none of the tested ligands present during receptor expression significantly stabilized the wt MOR during the immobilization (Figure 4B).

Ligand-binding characteristics of immobilized receptor variants

In order to assess fully the effect of solubilization in CHAPS–CHS on the functionality of immobilized receptor, we determined the ligand binding affinities of the immobilized receptor variants wt KOR, FCL KOR and chimera KKM to various ligands. Remarkably, for all receptor variants, there was no substantial difference in the ligand binding profile between the membrane embedded and the immobilized forms (Table I). Taken together, these results suggest that the receptor variants

wt KOR, FCL KOR and the chimera KKM are functional in all respects tested after immobilization in detergent-solubilized form.

Discussion

Stability engineering for GPCRs

KOR and MOR are highly homologous (about 60% identity), with the highest divergence occurring at the N-terminus of the receptors. However, we observed significant differences in their stability in detergent-solubilized form. Receptor chimeras have been used to elucidate sequence–function relationships for many GPCRs; however, the stability in detergent has not been investigated.

In the present study, we constructed the chimera KKM, where the N-terminus up to the first cytoloop originates from wt KOR and the rest of the sequence is taken from wt MOR (Figure 1), and we observed that KKM shows substantially increased expression levels and enhanced stability in detergent when compared with MOR. The ligand binding profiles of immobilized and membrane-embedded KKM do not differ (see above) and the maintenance of ligand binding affinities was also found for the immobilized KOR. Our study points to a yet unexplored receptor region for possible GPCR engineering. Based on the vast amount of mutagenesis data for opioid receptors (Bot *et al*., 1998; Lu *et al*., 1998; Law *et al*., 1999; Mollereau *et al*., 1999; Ide *et al*., 2000; Feng *et al*., 2001; Scearce-Levie *et al*., 2001; Decaillot *et al*., 2003; Tanowitz and von Zastrow, 2003; Wang *et al*., 2003) and on GPCR structural models (Pogozheva *et al*., 1998; Chavkin *et al*., 2001; McFadyen *et al*., 2002; Visiers *et al*., 2002; Archer *et al*., 2003; Mirzadegan *et al*., 2003), the receptor N-terminal region does not seem to be directly involved in the ligand binding and/or interaction with G-proteins; however, its alteration might result in significantly increased expression and/or stability of the receptor, for instance in detergents. As the chimeras with either only the N-terminal region or only TM1 taken from KOR into MOR did not improve these properties, we have as yet no molecular-level explanation for the beneficial effect of the KKM mutant. Receptor expression can be influenced by the translocation of the N-terminal tail (Andersson *et al*., 2003), which is longer in MOR than in KOR (Figure 1), and one may speculate that the N-terminal region in such receptors, which do not have a cleavable signal sequence, affects their membrane insertion and/or degradation. However, additional effects within the protein must mediate the greater robustness of KKM in detergent after solubilization, especially when compared with MOR. The exchange of only the extracellular N-terminal region did not improve the properties, possibly since the point of junction we chose may have led to molecules with incorrect helical length for both chimeras KMM and MKM. Furthermore, we have engineered an N-terminal FLAG tag in front of all receptor variants studied here. It is possible that this tag enhances the intrinsic difference in the effect of N-terminal regions on the receptor expression. The most important result is, however, that KKM can be immobilized in functional form with ~20-fold higher yield than wt MOR under the same conditions (Figure 4A). This is certainly encouraging for further GPCR engineering.

We do not yet know the true state of the receptor immobilized on the beads. Especially, we cannot state whether the functional, immobilized receptor is monomeric or oligomeric. Even though detergent concentrations above the CMC were used, we also cannot present direct evidence that the immobilized receptor is exclusively in a micellar state. However, the lack of activity of MOR and some other receptor variants in the immobilized state may actually be taken as supporting evidence that the immobilized receptor is probably not embedded in small patches of remaining lipid bilayers. All tested receptor variants (wt MOR, KMM, MKM, KKM, wt KOR, KOR FCL) were clearly functional in the cell or in membrane preparations.

There are reports on the construction of chimeras which resemble our chimeras. However, the stability in detergents has not been investigated for any of them. In two studies, the whole N-terminal region including the first TM helix originated from the δ-opioid receptor (DOR) and the rest of the sequence from MOR and the ligand-binding profile of the chimera is almost identical with the profile of wt MOR (Onogi *et al*., 1995; Wang *et al*., 1995). In another report, two chimeras were constructed where again the whole N-terminal region including the first TM helix originated either from DOR or from KOR and the rest of sequence from MOR (Ide *et al*., 2000). No significant deviation from the ligand-binding profile of the wt MOR was observed for the two chimeras. These reports, together with the present study (Table I), indicate that the protein sequence of the N-terminal part of the MOR is not important for the specificity of ligand binding for the ligands under study.

Solubilization and immobilization of opioid receptors

We also present here an optimized immobilization procedure, which allows biophysical characterizations of opioid receptors in a highly enriched and active form. The investigation of ligand binding to an immobilized receptor has many advantages over other binding assays with solubilized receptors in solution, which require the subsequent separation of bound from unbound ligand. Although this is easily achieved with filtration in the case of membrane fragments, it is much less straightforward in the case of solubilized receptor. Therefore, it had not been easy to troubleshoot the lack of binding signal for unknown receptors with such assays, as this could come either from loss of receptor activity upon solubilization or problems with filter-binding assays. However, the immobilization of solubilized receptors to magnetic beads circumvents this problem and makes it convenient to handle the solubilized receptor for binding assays and thus allows one to test the influence of detergents and other factors on stability.

Most GPCRs are not stable in detergents, but the extraction of receptor from the membrane by using detergent is a prerequisite for any biophysical study of the receptor structure. The solubilized and purified MOR had been reported to have significantly reduced (~1000-fold) affinity for opioid agonists (Simon, 1991; Ofri *et al*., 1992; Gioannini *et al*., 1993). The exact molecular reasons for the loss of GPCR function in detergents are not known and have been attributed to a dissociation of lipids (Lagane *et al*., 2000; Garavito and Ferguson-Miller, 2001; Lee, 2003) and/or G-proteins (Ofri *et al*., 1992; Fan *et al*., 1995; Stanasila *et al*., 1999) from specific sites on the receptor or to the disassembly of receptor oligomers in the presence of detergent or to receptor denaturation and aggregation because of water access to the receptor in detergent micelles. Protein immobilized on a solid phase is much better protected from aggregation (Li *et al*., 2004). We decided to

immobilize opioid receptors on paramagnetic beads, as they were successfully used previously for the efficient immobilization of other GPCRs, such as chemokine receptors CCR5 (Mirzabekov *et al*., 2000) or CXCR4 (Babcock *et al*., 2001) in an active form. In this case, however, supported lipid bilayers were used, rather than the detergent-solubilized form used in our study.

There are reports about immobilization of opioid receptors in functional form on a protein A-coated Sepharose resin by using an anti-MOR antibody (Chalecka-Franaszek *et al*., 2000) or on an artificial phospholipid-monolayer support (Beigi and Wainer, 2003); however, evidence that the receptors were completely extracted from the lipid bilayers prior or during the immobilization was not presented.

We have also shown here that it was possible to immobilize the KOR devoid of all free cysteines (FCL receptor) in functional form. The KOR FCL had been designed for mapping of the conformational changes in KOR molecule by using fluorescence microscopy (Ott *et al*., 2004).

Our optimized method for the oriented immobilization of highly enriched opioid receptors can find application not only in studies of receptor conformational changes (by using site-specific labeled receptor and fluorescence microscopy) and of G-protein coupling (by using surface plasmon resonance), but also especially for testing of various GPCR mutants and conditions under which the detergent-solubilized protein can be retained functional. It also allows drug identification in high-throughput screening. Selection from protein libraries against the immobilized receptor (Ostermeier *et al*., 1995; Padan *et al*., 1998; Grisshammer *et al*., 2002; Röthlisberger *et al*., 2004) may provide stable native-receptor-conformation specific binders, which might further stabilize the receptor during its biophysical and/or structural characterization.

Acknowledgements

The authors thank Drs Stephen F.Marino, Tomoaki Matsuura and Manca Kenig for valuable discussions. This work was supported by the National Center for Competence in Research in Structural Biology and by TopNano21 Project (KTI, Berne, Switzerland).

References

Andersson,H., D'Antona,A.M., Kendall,D.A., Von Heijne,G. and Chin,C.N. (2003) *Mol. Pharmacol.*, **64**, 570–577.
Archer,E., Maigret,B., Escrieut,C., Pradayrol,L. and Fourmy,D. (2003) *Trends Pharmacol. Sci.*, **24**, 36–40.
Babcock,G.J., Mirzabekov,T., Wojtowicz,W. and Sodroski,J. (2001) *J. Biol. Chem.*, **276**, 38433–38440.
Banères,J.L. and Parello,J. (2003) *J. Mol. Biol.*, **329**, 815–829.
Banères,J.L., Martin,A., Hullot,P., Girard,J.P., Rossi,J.C. and Parello,J. (2003) *J. Mol. Biol.*, **329**, 801–814.
Beigi,F. and Wainer,I.W. (2003) *Anal. Chem.*, **75**, 4480–4485.
Bot,G., Blake,A.D., Li,S. and Reisine,T. (1998) *J. Neurochem.*, **70**, 358–365.
Chalecka-Franaszek,E., Weems,H.B., Crowder,A.T., Cox,B.M. and Cote,T.E. (2000) *J. Neurochem.*, **74**, 1068–1078.
Chavkin,C., McLaughlin,J.P. and Celver,J.P. (2001) *Mol. Pharmacol.*, **60**, 20–25.
Clark,W.A., Jian,X., Chen,L. and Northup,J.K. (2001) *Biochem. J.*, **358**, 389–397.
Conway,B.R. and Demarest,K.T. (2002) *Receptors Channels*, **8**, 331–341.
Decaillot,F.M., Befort,K., Filliol,D., Yue,S., Walker,P. and Kieffer,B.L. (2003) *Nat. Struct. Biol.*, **10**, 629–636.
Doi,T., Hiroaki,Y., Arimoto,I., Fujiyoshi,Y., Okamoto,T., Satoh,M. and Furuichi,Y. (1997) *Eur. J. Biochem.*, **248**, 139–148.
Fan,L.Q., Gioannini,T.L., Wolinsky,T., Hiller,J.M. and Simon,E.J. (1995) *J. Neurochem.*, **65**, 2537–2542.
Feng,Y.P., Chen,L.W., Zhou,D.H., Chen,J., Xu,X.J. and Chi,Z.Q. (2001) *Acta Pharmacol. Sin.*, **22**, 981–985.
Garavito,R.M. and Ferguson-Miller,S. (2001) *J. Biol. Chem.*, **276**, 32403–32406.
Gaveriaux-Ruff,C. and Kieffer,B.L. (2002) *Neuropeptides*, **36**, 62–71.
Gioannini,T.L., Fan,L.Q., Hyde,L., Ofri,D., Yao,Y.H., Hiller,J.M. and Simon,E.J. (1993) *Biochem. Biophys. Res. Commun.*, **194**, 901–908.
Grisshammer,R., Grunwald,T. and Sohal,A.K. (2002) *Protein Expr. Purif.*, **24**, 505–512.
Ide,S., Sakano,K., Seki,T., Awamura,S., Minami,M. and Satoh,M. (2000) *Jpn. J. Pharmacol.*, **83**, 306–311.
Karlsson,O.P. and Löfas,S. (2002) *Anal. Biochem.*, **300**, 132–138.
Kassack,M.U., Hofgen,B., Lehmann,J., Eckstein,N., Quillan,J.M. and Sadee,W. (2002) *J. Biomol. Screening*, **7**, 233–246.
Lagane,B., Gaibelet,G., Meilhoc,E., Masson,J.M., Cezanne,L. and Lopez,A. (2000) *J. Biol. Chem.*, **275**, 33197–33200.
Law,P.Y., Wong,Y.H. and Loh,H.H. (1999) *Biopolymers*, **51**, 440–455.
Lee,A.G. (2003) *Biochim. Biophys. Acta*, **1612**, 1–40.
Li,J., Chen,C., Huang,P. and Liu-Chen,L.Y. (2001) *Mol. Pharmacol.*, **60**, 1064–1075.
Li,L.Y., Su,Y.F., Zhang,Z.M., Wong,C.S. and Chang,K.J. (1993) *NIDA Res. Monogr.*, **134**, 146–164.
Li,M., Su,Z.G. and Janson,J.C. (2004) *Protein Expr. Purif.*, **33**, 1–10.
Lu,Y.F. *et al.* (1998) *Synapse*, **28**, 117–124.
McFadyen,I., Metzger,T., Subramanian,G., Poda,G., Jorvig,E. and Ferguson,D.M. (2002) *Prog. Med. Chem.*, **40**, 107–135.
McLean,A.J., Zeng,F.Y., Behan,D., Chalmers,D. and Milligan,G. (2002) *Mol. Pharmacol.*, **62**, 747–755.
Mirzabekov,T., Kontos,H., Farzan,M., Marasco,W. and Sodroski,J. (2000) *Nat. Biotechnol.*, **18**, 649–654.
Mirzadegan,T., Benko,G., Filipek,S. and Palczewski,K. (2003) *Biochemistry*, **42**, 2759–2767.
Mollereau,C., Mouledous,L., Lapalu,S., Cambois,G., Moisand,C., Butour,J.L. and Meunier,J.C. (1999) *Mol. Pharmacol.*, **55**, 324–331.
Neumann,L., Wohland,T., Whelan,R.J., Zare,R.N. and Kobilka,B.K. (2002) *ChemBioChem*, **3**, 993–998.
Ofri,D., Ritter,A.M., Liu,Y.F., Gioannini,T.L., Hiller,J.M. and Simon,E.J. (1992) *J. Neurochem.*, **58**, 628–635.
Onogi,T., Minami,M., Katao,Y., Nakagawa,T., Aoki,Y., Toya,T., Katsumata,S. and Satoh,M. (1995) *FEBS Lett.*, **357**, 93–97.
Ostermeier,C., Iwata,S., Ludwig,B. and Michel,H. (1995) *Nat. Struct. Biol.*, **2**, 842–846.
Ott,D., Frischknecht,R. and Plückthun,A. (2004) *Protein Eng. Des. Sel.*, **17**, 37–48.
Padan,E., Venturi,M., Michel,H. and Hunte,C. (1998) *FEBS Lett.*, **441**, 53–58.
Palczewski,K. *et al.* (2000) *Science*, **289**, 739–745.
Pogozheva,I.D., Lomize,A.L. and Mosberg,H.I. (1998) *Biophys. J.*, **75**, 612–634.
Reece,P.A., Sedman,A.J., Rose,S., Wright,D.S., Dawkins,R. and Rajagopalan,R. (1994) *J. Clin. Pharmacol.*, **34**, 1126–1132.
Röthlisberger,D., Pos,K.M. and Plückthun,A. (2004) *FEBS Lett.*, **564**, 340–348.
Scearce-Levie,K., Coward,P., Redfern,C.H. and Conklin,B.R. (2001) *Trends Pharmacol. Sci.*, **22**, 414–420.
Simon,E.J. (1986) *Ann. N. Y. Acad. Sci.*, **463**, 31–45.
Simon,E.J. (1991) *Med. Res. Rev.*, **11**, 357–374.
Smith,A.P. and Loh,H.H. (1991) *NIDA Res. Monogr.*, **111**, 69–84.
Stanasila,L., Massotte,D., Kieffer,B.L. and Pattus,F. (1999) *Eur. J. Biochem.*, **260**, 430–438.
Stenlund,P., Babcock,G.J., Sodroski,J. and Myszka,D.G. (2003) *Anal. Biochem.*, **316**, 243–250.
Tanowitz,M. and von Zastrow,M. (2003) *J. Biol. Chem.*, **278**, 45978–45986.
Tucker,J. and Grisshammer,R. (1996) *Biochem. J.*, **317**, 891–899.
Visiers,I., Ballesteros,J.A. and Weinstein,H. (2002) *Methods Enzymol.*, **343**, 329–371.
Wang,W., Loh,H.H. and Law,P.Y. (2003) *J. Biol. Chem.*, **278**, 36848–36858.
Wang,W.W., Shahrestanifar,M., Jin,J. and Howells,R.D. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 12436–12440.
Weems,H.B., Chalecka-Franaszek,E. and Cote,T.E. (1996) *J. Neurochem.*, **66**, 1042–1050.
Wise,A., Jupe,S.C. and Rees,S. (2004) *Annu. Rev. Pharmacol. Toxicol.*, **44**, 43–66.
Zukin,R.S. and Maneckjee,R. (1986) *Methods Enzymol.*, **124**, 172–190.

Received February 15, 2005; accepted February 18, 2005

Edited by David Eisenberg

Chapter 4

Directed evolution of a G-protein coupled receptor for expression, stability, and binding selectivity

4.1 Published article

Directed evolution of a G protein-coupled receptor for expression, stability, and binding selectivity

Casim A. Sarkar*†‡, Igor Dodevski*†, Manca Kenig*§, Stefan Dudli*, Anja Mohr*, Emmanuel Hermans¶, and Andreas Plückthun*‖

*Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland; and ¶Laboratoire de Pharmacologie Expérimentale, Université catholique de Louvain, Avenue Hippocrate 54.10, B-1200 Bruxelles, Belgium.

Edited by William F. DeGrado, University of Pennsylvania, Philadelphia, PA, and approved July 18, 2008 (received for review April 1, 2008)

We outline a powerful method for the directed evolution of integral membrane proteins in the inner membrane of *Escherichia coli*. For a mammalian G protein-coupled receptor, we arrived at a sequence with an order-of-magnitude increase in functional expression that still retains the biochemical properties of wild type. This mutant also shows enhanced heterologous expression in eukaryotes (12-fold in *Pichia pastoris* and 3-fold in HEK293T cells) and greater stability when solubilized and purified, indicating that the biophysical properties of the protein had been under the pressure of selection. These improvements arise from multiple small contributions, which would be difficult to assemble by rational design. In a second screen, we rapidly pinpointed a single amino acid substitution in wild type that abolishes antagonist binding while retaining agonist-binding affinity. These approaches may alleviate existing bottlenecks in structural studies of these targets by providing sufficient quantities of stable variants in defined conformational states.

integral membrane proteins | protein engineering | protein folding

G protein-coupled receptors (GPCRs) comprise ≈1% of the genes in mammalian genomes and constitute ≈60% of all drug targets (1). Given the critical importance of this class of integral membrane proteins, there is great interest in having detailed structures of these molecules. However, in the Protein Data Bank (2), which contains ≈18,000 nonredundant protein structures, the structures of only two GPCRs have been deposited, that of bovine rhodopsin (3), uniquely facilitated by its natural abundance in the retina (4), and very recently that of β2 adrenergic receptor (5), whose high resolution structure determination required the replacement of a loop by a whole protein (6).

Because most GPCRs are normally produced at very low levels, an overexpression system must be set up for each GPCR of interest to obtain quantities sufficient for biophysical analyses. Many attempts have been made to increase functional membrane protein expression levels by testing various combinations of variables defining the expression system, such as host organism, expression plasmid, fusion adducts, codon usage, and expression induction conditions (7). However, even if all of these parameters could be globally optimized, they still may not yield enough functional material because of inherent limitations in the protein sequence itself. Furthermore, the limited stability of solubilized GPCRs is another bottleneck for their biophysical investigation, and this cannot be influenced by the expression system. Thus, while it may be possible to solve a limited number of new GPCR structures whose sequences are sufficiently accommodating to support such a brute-force screening for expression and solubilization conditions, this method is unlikely to have general applicability in routinely generating large amounts of stable, solubilized protein for any given GPCR.

To directly address the importance of receptor sequence as an experimental variable in membrane protein expression and stability, we have developed a powerful approach, inspired by periplasmic expression with cytometric screening (8) and anchored periplasmic expression (9), in which we evolve the sequence of a GPCR, keeping all other variables constant, to yield more functionally expressed protein in a convenient heterologous host, *Escherichia coli*. We used as a model system the rat neurotensin receptor-1 (NTR1), which has been shown to give a detectable yield in *E. coli* (10, 11) but which still needs to be improved to allow more convenient preparation of milligram quantities of receptor.

Detailed characterization of the best variant from the selection reported here reveals that it exhibits an order-of-magnitude increase in expression level in both *E. coli* and *Pichia pastoris*, elevated expression in mammalian cells, and enhanced stability in detergent-solubilized form, yet it largely retains the biochemical properties of WT NTR1, including binding affinity, binding selectivity, and G protein-mediated signaling. We have further extended this approach to isolate mutants of NTR1 with altered ligand selectivity. This methodology should thus be of general utility in the directed evolution of stable variants of such proteins to high-level expression in multiple states of activity.

Results and Discussion

Setup and Optimization of Screening Methodology. The general approach is given in Fig. 1. The expression vector containing the GPCR library of interest (e.g., from an error-prone PCR of the receptor gene) with two constant fusion partners (N-terminal maltose binding protein and C-terminal thioredoxin) is used to express the corresponding proteins in functional form in the inner membrane of *E. coli* DH5α (see *Methods*). After expression, cells are incubated at 4°C in an optimized buffer that renders the outer membrane permeable to small molecules to allow binding of fluorescent ligand to the receptors, and at the same time maximizes cell viability (see *Methods*). Chen and colleagues (8) have previously described conditions in which ligands as large as 10 kDa can enter the periplasmic space of *E. coli* without compromising cell viability. Here, the buffer was specifically optimized to allow for saturable, specific binding of fluorescently labeled agonist [BODIPY-NT(8–13)] to NTR1, while maximizing cell viability after fluorescence-activated cell sorting (FACS) [see supporting information (SI) Figs. S1–S4]. Although the expression conditions for WT NTR1 typically

Author contributions: C.A.S., I.D., E.H., and A.P. designed research; C.A.S., I.D., M.K., S.D., and A.M. performed research; E.H. contributed new reagents/analytic tools; C.A.S., I.D., M.K., A.M., E.H., and A.P. analyzed data; and C.A.S., I.D., M.K., A.M., E.H., and A.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

†C.A.S. and I.D. contributed equally to this work.

‡Present address: Departments of Bioengineering and Chemical and Biomolecular Engineering, University of Pennsylvania, 240 Skirkanich Hall, 210 South 33rd Street, Philadelphia, PA 19104-6321.

§Present address: Lek Pharmaceuticals, Kolodvorska 27, SI-1234 Menges, Slovenia.

‖To whom correspondence should be addressed. E-mail: plueckthun@bioc.uzh.ch.

This article contains supporting information online at www.pnas.org/cgi/content/full/0803103105/DCSupplemental.

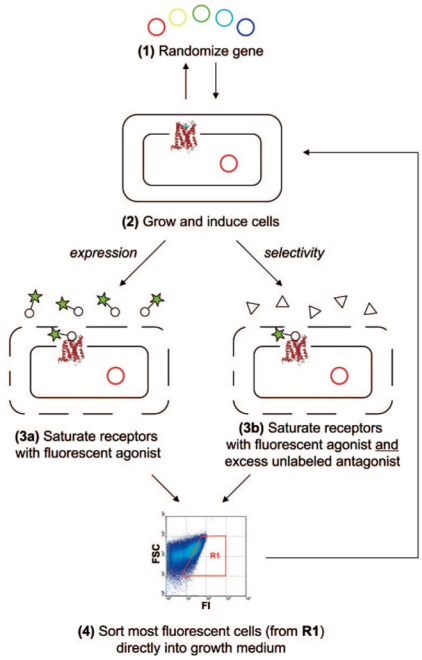


Fig. 1. General selection scheme for increasing expression level (steps 1, 2, 3a, 4, back to 2) and altering ligand selectivity (steps 1, 2, 3b, 4, back to 2).

generate <1,000 functional receptors per cell (10, 11), optimization of the binding buffer and FACS gating conditions resulted in a specific signal in the gating window that was ≈900-fold above background (see Fig. S4).

After incubation with saturating concentrations of BODIPY-NT(8–13), bacteria expressing the largest number of functional receptors correspondingly exhibit the greatest fluorescence, and these cells were collected directly in growth medium and then expanded for a subsequent round. A single selection round, which consisted of library expansion, induced receptor expression, incubation with fluorescent ligand, and FACS to recover the most fluorescent bacterial cells, took approximately 1 day. The advantage of maintaining viable cells was that they could be immediately regrown after sorting, thus eliminating any preparative steps between selection rounds. Whenever additional diversity was desired after any FACS round, the sorted pool of cells was grown and harvested, the enriched plasmid collection was purified, the GPCR sequences (excluding the fusion partners) were further randomized, and fresh bacteria were transformed for the next selection. The flowchart for the selections on NTR1 is given in Fig. S5.

Table 2. K_D of full-length neurotensin binding to receptors on whole cells

	NTR1	D03
E. coli	0.14 ± 0.01 nM	0.11 ± 0.01 nM
HEK293T	3.7 ± 0.5 nM	1.9 ± 0.2 nM

Selection of Variants with Increased Expression Level. For increasing expression level, the initial randomized NTR1 library was subjected to four rounds of FACS. In each round, only the most fluorescent ≈0.1 to 1% of the cells were collected. Nonetheless, after these rounds, the evolved pool had a mean fluorescence intensity (MFI) no greater than that of the WT sequence. Error-prone PCR (epPCR) was used to overlay another set of random mutations on top of those that were enriched after the first four rounds of FACS, and this rerandomized library was again subjected to four rounds of sorting. In this second set of sorts, the MFI of the pool overtook that of WT NTR1. After a third randomization step followed by four more rounds of FACS, the evolved pool was split into two. One half was randomized by epPCR a fourth time and the other half was shuffled, using the staggered extension process (StEP) (12).

After these selections, the MFI was approximately five times that of WT NTR1. From the enriched pool, 96 single clones were sequenced and analyzed for receptor expression level (see *Methods* and Figs. S6 and S7). The clone with the best functional receptor expression level per cell, D03, exhibited approximately a 10-fold increase in specific signal, as assayed by [^{3}H]-NT binding and flow cytometry (Table 1 and Fig. S8). D03 has 14 nucleotide substitutions scattered throughout its helices and loops (see Fig. S10 for a snake-like plot with the exact mutational positions). Five of these mutations are silent, suggesting that incorporation of nonsilent mutations was slow, approximately two amino acid substitutions per round of epPCR. This may be due to the seven-transmembrane helical topology of the protein, which limits the number and type of mutations that are possible, mandating a strategy with low mutational load.

Evolved Receptor Retains Biochemical and Pharmacological Properties of WT. In the membrane of both *E. coli* and mammalian cells, D03 binds NT with affinities comparable to WT (Table 2), as determined by radioligand binding assays (see *Methods*). The radioactive signal is almost entirely competed away with a 50-fold excess of either unlabeled NT (agonist) or SR 48692 (antagonist) (13), suggesting that binding specificity is also faithfully retained (Fig. 2).

We also compared the signaling properties of the evolved mutant to those of WT. NTR1 signals mainly via the $G_{q/11}$ subtype of G proteins (14), which triggers the mobilization of intracellular Ca^{2+} pools ($[Ca^{2+}]_i$) via phospholipase C (PLC)-generated IP_3. The use of a fluorescent Ca^{2+} indicator, such as Fura 2, enables detection of variations in $[Ca^{2+}]_i$ upon agonist binding to NTR1 (see *Methods*). Because of the different types

Table 1. Expression levels of NTR1, D03, and D03-L167R in multiple hosts

	NTR1	D03 (fold of NTR1)*	D03-L167R (fold of NTR1)*
E. coli,† no. per cell	705 ± 101	6,155 ± 777‡ (8.7 ± 1.7)	4,647 ± 879 (6.6 ± 1.6)
P. pastoris,§ pmol/mg	18.3 ± 5.4	216.7 ± 36.9 (11.9 ± 4.1)	51.1 ± 4.3 (2.8 ± 0.9)
HEK293T, no. per cell	18,300 ± 1,300	58,800 ± 6,600 (3.2 ± 0.4)	33,300 ± 5,400 (1.8 ± 0.3)

*Parenthetical terms give the ratio of mutant expression to that of NTR1.
†Expression in 1 l cultures.
‡Expression in mutated vector pRG_{C7054G}.
§Functional receptor (pmol) normalized to total membrane protein content (mg). NT binding cannot be measured with whole *P. pastoris* cells.

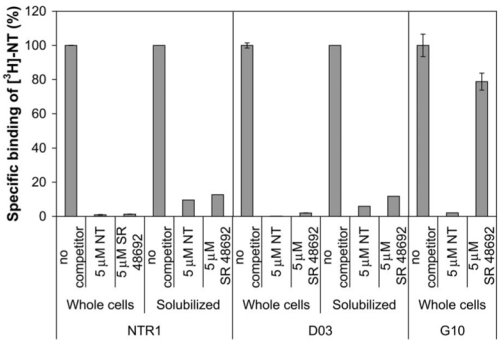


Fig. 2. Percentage radioligand binding to NTR1, D03 (evolved expression variant), and G10 (evolved selectivity variant) with unlabeled agonist (NT) and antagonist (SR 48692) competitors (used at 5 μM each). The radioligand was used at 10 nM. D03 retains the ligand binding properties of NTR1, whereas G10, which was evolved under selection pressure to abolish binding to SR 48692, no longer exhibits high affinity to antagonist.

of Ca^{2+} signaling patterns (oscillations, plateau, and transient; see Fig. S11), experiments with pooled cells cannot easily be evaluated, and we thus performed single-cell measurements of Ca^{2+} signaling in HEK293T cells transiently transfected with either WT or D03. It is well known that mutations in the highly conserved (D/E)R(W/Y) motif in helix III of Class A GPCRs can affect ligand binding and G protein coupling (15). D03 contains a substitution of Arg^{167} with Leu in this motif; this does not appear to influence the binding of NT to D03 (see Table 2), yet could potentially influence signaling. To check this, we also constructed a back mutant, D03-L167R, in which this important three amino acid motif was restored.

For WT NTR1, oscillations (the response at the lowest agonist concentration) were detected at 10^{-11} M NT, while for D03, these were only observed with a 100-fold higher concentration of agonist (Fig. 3), suggesting that the receptor can indeed still signal but that its coupling efficiency to $G_{q/11}$ is reduced. By contrast, the single amino acid revertant, D03-L167R, was capable of generating oscillatory behavior at the same minimal NT concentration (10^{-11} M) as WT. The concentration of NT (1 nM) needed to achieve strong transient and plateau responses (no oscillations) is the same for both WT and D03-L167R (see Fig. 3), while >100 nM agonist is needed for D03. For NTR1 and D03-L167R, the minimum concentration of agonist that leads to significant transient and plateau responses (> 40%) is similar to

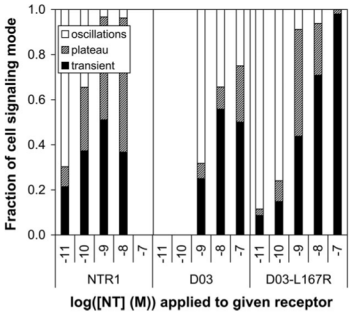


Fig. 3. Ca^{2+} signaling in HEK293T cells expressing NTR1, D03, or D03-L167R. Single-cell measurements were performed and the resulting signal was classified as an oscillatory, plateau, or transient response (see *Methods*, *SI Text*, and Fig. S11).

that at which the oscillatory response disappears, also approximately 1 nM, while it is again >100 nM for D03. These findings underscore the importance of the (D/E)R(W/Y) motif in G protein coupling (15). While these results have been obtained with constructs missing amino acids 1 to 42 of the GPCR, no major difference in signaling between the truncated and full-length forms was detected (see Fig. S12).

That these evolved receptors can qualitatively mimic the signals generated by WT, including similar agonist concentration-dependent signaling of D03-L167R, is pleasantly surprising for two reasons: (*i*) the variants are significantly mutated (8–9 amino acid substitutions) and (*ii*) the ability to efficiently signal was never under the pressure of selection. More generally, this suggests that it is indeed possible to partially decouple the biophysical properties of a GPCR from its biochemical and pharmacological properties through mutagenesis, which thus enables the use of protein engineering approaches such as the present methodology to improve the robustness of these membrane proteins for structural studies.

Evolved Receptor Shows Increased Expression in Both Prokaryotic and Eukaryotic Hosts. There are at least two conceivable mechanisms by which more functional D03 is obtained in *E. coli* as a result of selection: (*i*) the total amount of NTR1 and D03 per cell is comparable, but the fraction of properly folded and inserted D03 is significantly greater, or (*ii*) the total amount of D03 per cell is significantly greater than that of NTR1, but the fraction of functional receptor is similar. Whole-cell Western blots of *E. coli* proteins reveal that more D03 molecules are detected per cell (see Fig. S13). While this would be consistent with a greater rate of biosynthesis of D03, it more likely reflects the fact that noninserted and nonfunctional WT is degraded. Thus, the quantity of properly inserted, functional GPCR seems to correlate well with the total amount of receptor detected.

To determine whether D03 may have merely adapted to the biosynthetic pathway or the membrane of *E. coli* during selection or may have actually acquired traits of generally improved biophysical properties, we also expressed WT NTR1, D03, and D03-L167R in the methylotrophic yeast *P. pastoris* and compared functional and total expression yields. The functional expression level for D03, as assayed by specific radioligand binding to membrane preparations, was ≈12-fold higher than that for WT NTR1 (Table 1), while the improvement for the backmutated D03-L167R was more modest. This is in contrast to *E. coli*, where the effect of the back mutation was very small (Table 1). When total receptor protein levels in *P. pastoris* are compared by Western blot (Fig. 4), the increased expression of D03 is confirmed by the strong intensity of the band at ≈43 kDa (see *Methods*). In contrast, WT NTR1 shows a strong band that has not properly entered the gel, suggesting that this receptor is more aggregation-prone than the evolved D03 and D03-L167R. We found that the precursor form [unprocessed prepro- or pro-alpha factor fusion (16)] is detected at about equal intensity for WT and mutants (see Fig. S14), suggesting that protein biosynthesis is not changed, but that the expression level in *P. pastoris* is determined by a folding step after translocation and processing, further supporting the hypothesis that the biophysical properties have been improved.

To test whether this improvement is also seen in mammalian cells, we transiently transfected HEK293T cells with NTR1 or D03. The evolved mutant was approximately threefold better expressed, as measured in whole-cell radioligand binding assays (see Table 1). Again, the effect of the back mutation L167R on expression is small (Table 1). Similar experiments in CHO cells and COS cells also revealed a two- to threefold increase in functional D03 expression relative to WT NTR1 (data not shown). In all of the experimentally tested hosts, neither the growth rate nor the final cell density was detectably different when expressing NTR1 and D03 in parallel cultures (data not shown). These results thus suggest that D03 is

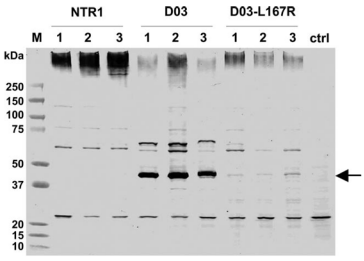


Fig. 4. Comparison of total GPCR yields in *P. pastoris* by immunoblot analysis. The C-terminal biotinylation was detected with a streptavidin-alkaline phosphatase conjugate. Three independent clones of each construct and a clone with an ''empty'' vector control (*ctrl*) integrated into the *P. pastoris* genome were analyzed (20 μg total membrane protein loaded per lane). Full-length, processed GPCR should run at the position of the arrow. The ≈60 kDa band present in all receptor samples represents the unprocessed precursor form (see Fig. S14). The ≈65 kDa band, which is more pronounced in the D03 samples, also reacts with both M1 and M2 anti-Flag antibodies (data not shown) and likely represents a compact dimeric form. The high molecular weight smear appears to be aggregated receptor, as it is diminished in D03 and D03-L167R. The band at ≈23 kDa, which also appears in the empty vector control, is a *P. pastoris* protein.

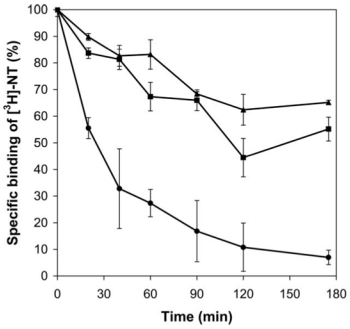


Fig. 5. Kinetics of thermal denaturation of purified NTR1 (●), D03 (▲), and D03-L167R (▪) as fusion proteins with maltose binding protein and thioredoxin. Equal amounts of each receptor were incubated at 45°C and specific activities were measured at various time points by saturation binding assays (see *Methods*). While NTR1 retains only ≈7% of its specific binding activity after 175 min, the evolved receptors D03 and D03-L167R retain ≈65% and ≈55% of their respective activities after the same interval of time. Results are duplicate measurements from a representative experiment.

an intrinsically more robust protein in expressing and folding not only in *E. coli*, but also in multiple eukaryotic hosts. It should be noted that all receptors were expressed as fusion proteins with N-terminal maltose binding protein (MBP) and C-terminal thioredoxin (TrxA) fusions in *E. coli*, whereas these fusion partners were not present for expression in eukaryotic hosts (*P. pastoris* and HEK293T cells). Therefore, the similar trends in functional expression (Table 1) and biochemical affinity (Table 2) across different hosts suggest that the fusion partners used in bacteria are not involved in causing these evolutionary improvements.

Characterization of Solubilized and Purified Receptor. To ascertain whether the observed enhancement in functional receptor expression in *E. coli* translates into a corresponding increase in pure, soluble material, we performed a side-by-side purification of the NTR1 and D03 fusion proteins (see *Methods*). Briefly, after detergent solubilization and purification by immobilized metal ion affinity chromatography (IMAC), the amount of functional GPCR was monitored at each step by a radiolabeled ligand-binding assay. As summarized in Table 3, the total amount of functional, purified D03 protein recovered from a 1 l expression is ≈6-fold that obtained for NTR1. A comparison of the gel filtration profiles for WT and D03 reveals almost identical elution behavior (see Figs. S17 and S18).

We then investigated whether the evolved receptors were also more stable in solubilized form. Solubilized and purified NTR1, D03, and D03-L167R were incubated at 45°C and the remaining activity was measured, after cooling, at various time points (Fig. 5). D03 and D03-L167R (in which the highly conserved DRY motif is restored) were both found to be significantly more thermostable than NTR1, suggesting that evolution with selection for increased functional expression favored proteins with improved biophysical properties. This implicit correlation between stability and functional expression level has been noted by others as well (17); thus, the present platform may serve to generate more stable variants of a given receptor.

Table 3. Purification of NTR1 and D03

	NTR1, pmol	D03, pmol	D03/NTR1*
Whole cell material	2,300	15,800	6.9
Solubilized material	1,700 ± 80	11,500 ± 1250	6.8
IMAC eluate	950 ± 5	6,020 ± 175	6.3

*Ratio of purification yields.

Tracing the Effects of Single Substitutions. Why is D03 better expressed and more stable? To understand the contribution of each of the mutations in D03 in enhancing expression level, a systematic, two-pronged site-directed mutagenesis strategy was used: (*i*) each of the substitutions in D03 was individually introduced into the WT sequence, and (*ii*) each of the existing substitutions in D03 was individually reverted back to the corresponding WT nucleotide. The first approach should elucidate which individual mutations can substantially increase the expression level on their own, whereas the second approach should reveal any additional mutations that also may be beneficial but whose effect may be masked by unfavorable WT amino acids. For the 28 variants that were generated in this manner, the receptor expression level was determined by saturation radioligand binding (see Fig. S19) and the impact of each single mutation—whether added to WT or subtracted from D03—is relatively small. We refer to this phenomenon as a ''staircase effect,'' because the improved phenotype of D03 is the sum of many incremental enhancements that leads to an increase in expression level.

In addition to the R167L mutation discussed previously, other unusual substitutions in NTR1 and one unintended mutation in the expression vector arose during the evolution of D03. Five of the nucleotide substitutions in NTR1 were silent, but two of these surprisingly introduced rare leucine codons (CTA) into the sequence (see Fig. S10). While it was recently shown that a synonymous single-nucleotide polymorphism in the *MDR1* gene can change the conformation of the corresponding protein because of altered kinetics of cotranslational folding and insertion (18), in our case none of the rare codons appears to alter protein expression (see Fig. S19) (although we cannot rigorously exclude that they might subtly influence protein conformation by altering biosynthesis kinetics). This result suggests that codon optimization or the use of rare tRNA-overexpressing strains may not be fruitful if other, inherently larger bottlenecks exist in heterologous protein expression.

Lastly, an unintentional mutation arose in the origin of replication in the plasmid harboring D03. This mutation (C7054G, for numbering see Fig. S15), close to the origin of replication and close to a previously identified mutation influ-

encing the copy number (19), increases the copy number of the plasmid by 100%, as determined by measuring the total amount of DNA from plasmid isolation experiments from standard expressions at 20°C (see Fig. S16). However, when D03 is expressed in the mutated pRG vector (pRG_{C7054G}), its expression level increases by only ≈25% as compared to the original vector (pRG_{WT}), and when WT NTR1 is expressed in the mutated vector, there is even no noticeable increase in its expression level (see Fig. S16). We hypothesize that the major bottlenecks in functional protein expression for WT NTR1 are cotranslational folding and insertion and, therefore, any other improvements along the expression pathway remain masked.

Selection of Variants with Altered Ligand Selectivity. For selections of ligand selectivity on NTR1 (see Fig. 1), two ligands were used, agonist BODIPY-NT(8–13) and antagonist SR 48692. In principle, selectivity selections could be performed independently of selections for expression level. However, in the case of NTR1, the FACS signal of the WT receptor was so weak with BODIPY-NT(8–13) alone, that the addition of excess SR 48692 would drop the signal of any positive clones into the background. Thus, the selectivity screen was only used after an initial selection for expression level to ensure that the MFI of the pool was significantly above background (after the fourth epPCR; see Fig. S5). Sequencing of the library before the fourth epPCR revealed expression-specific mutations; thus, any mutations that were enriched after the fourth epPCR in this screen could be readily identified as selectivity mutations and not expression mutations.

After performing selections for NTR1 sequences that still bound BODIPY-NT(8–13) in the presence of 100-fold excess SR 48692, 96 single clones were sequenced and analyzed for receptor expression level. The best expressing clone, G10, exhibited approximately a fivefold higher expression level than the WT, could be fully competed by NT, but could not be fully competed by 100-fold excess SR 48692 (see Fig. 2 and Fig. S9).

In stark contrast to the staircase effect observed with D03 and other expression variants, there was only one consensus mutation observed in G10 and other selectivity variants. This "elevator effect" in changing selectivity arose from a single mutation, F358S (see Fig. S10). The effects of mutation at Phe-358 have been studied by others (20), and reveal that a substitution to alanine at this position results not only in decreased antagonist affinity, but also in spontaneous basal inositol phosphate production in a receptor-dependent manner (21). Based upon a sequence alignment with bovine rhodopsin, Phe-358 may play an important role in maintaining the interaction between transmembrane helices 6 and 7, which in turn keeps NTR1 in an inactive conformation (21). Disruption of this residue leads to the observed constitutive activity. Thus, more generally, the present methodology has the ability to rapidly isolate mutations that may trap receptors in the active or inactive state, an approach that is complementary to previous work in engineering GPCR selectivity and activity (22, 23).

Conclusion

There have been recent advances in membrane protein engineering, including technologies for the screening of high-expressing members in a diverse pool of eukaryotic membrane proteins (24), identification of functionally critical amino acids in a GPCR (25–27), manual blot screening of randomly mutated membrane proteins for increased expression (28), and introduction of thermostabilizing mutations, individually identified by trial and error, in a GPCR (29). Complementary to such approaches, we present here a powerful, high-throughput platform for the directed evolution of a GPCR to enhance both expression level and stability while retaining function and to tailor ligand selectivity. This methodology should be applicable to other integral membrane proteins as well, for which specific binding ligands are available, and may help to facilitate biophysical studies by allowing milligram-level production of these proteins in multiple states of activity. Importantly, no basal heterologous expression of a WT GPCR sequence is necessary for the approach to work, as long as there exist expressing mutants that can be recovered by FACS. In the present study, a highly expressing, stable NTR1 variant displays WT biochemical properties—as assayed by binding affinity, binding selectivity, and G protein mediated signaling—and therefore provides a biologically meaningful template for structural studies. Such evolved membrane proteins may facilitate X-ray crystallography trials not only because they can be produced more abundantly but also because they remain functional in detergent micelles for significantly longer periods of time. If the proteins are more stable because of enhanced rigidity, they may also be more likely to generate properly diffracting crystals for structural determinations. The ability to obtain such high-resolution structures may help to elucidate the molecular basis for activation, inactivation, or pathology associated with that receptor, and may also provide templates for drug design.

Methods

Library Design and Selection. The rat neurotensin receptor-1 gene (NTR1; amino acids 43–424) has previously been expressed in *E. coli* using a vector that generates an N-terminal fusion of the receptor to MBP and C-terminal fusion to TrxA to enhance expression. The fusion protein contains tobacco etch virus (TEV) protease cleavage sites on both ends of the receptor and a C-terminal His_{10} tag. This derivative of the expression vector pRG/III-hs-MBP containing the NTR1 gene with these fusions was a kind gift from R. Grisshammer (National Institutes of Health). Expression of NTR1 in *E. coli* DH5α was essentially as described in ref. 30. Details of the preparation of fluorescently labeled neurotensin [BODIPY-NT(8–13)] and construction of NTR1 libraries are given in the *SI Text*. To allow binding of BODIPY-NT(8–13) to NTR1 in the inner membrane of *E. coli* while maximizing cell viability, an optimized binding buffer (50 mM Tris·HCl, pH 7.4, and 150 mM KCl) was identified (see *SI Text*). During each round of FACS, only the most fluorescent ≈0.1 to 1% of the cells in the population (≈10^7–10^8 cells) were recovered during sorting for regrowing and further selection. Individual cells from the final selections were sorted directly into 96-well plates during FACS and regrown to perform single clone analysis of expression levels by radioligand binding assays.

Radioligand Binding Assays. Details of the experimental protocols are given in the *SI Text*. Briefly, quantitative measurements of receptor number in *E. coli*, *P. pastoris*, HEK293T cells, and detergent solution were performed using a saturating concentration of radioactive agonist [^{3}H]-NT (10 nM) (PerkinElmer). For determining equilibrium binding affinities, a dilution series of radioligand was used (0.04–20 nM). Nonspecific binding was determined in the presence of 5 μM unlabeled NT. For competition experiments, antagonist SR 48692 (Sanofi Aventis) was used at a concentration of 5 μM.

Mammalian Cell Culture and Transfection. For expression in mammalian cells, receptors were cloned into the vector pcDNA3.1 (Invitrogen) encoding C-terminal Myc and His_6 tags. The HEK293T cell line (31) (a clonal line of HEK293 cells stably expressing SV40 large T antigen) was grown as described in ref. 32. Cells were routinely seeded into 6-well culture plates or 10 mm Petri dishes and grown for 24 h, reaching 70 to 80% confluence before transfection. Cells for binding assays were transiently transfected with DNA using calcium phosphate precipitation as described in ref. 32.

Single-Cell Monitoring of Variations in Intracellular Free Calcium, $[Ca^{2+}]_i$. Changes in $[Ca^{2+}]_i$ in response to NT were measured in individual HEK293T cells using the indicator Fura2-acetoxymethyl ester, as described in ref. 33. Different types of Ca^{2+} responses were observed and were classified as transient, oscillatory, or plateau (see *SI Text*).

Expression in *Pichia pastoris*. The *P. pastoris* strain SMD1163 (Invitrogen) was used for all experiments. A modified version of the yeast shuttle vector pPICZαC (Invitrogen) was previously designed in which the *Saccharomyces cerevisiae* alpha-factor prepro sequence (34), under the control of the AOX promoter, was followed by a Flag-M2 tag, a His_{10} tag, a TEV cleavage site, the GPCR, the linker sequence EFELGTRGS, and a biotin acceptor (BioAcc) domain (SwissProt P02904, amino acids 50–123). As a negative control in all experi-

ments, the plasmid without GPCR insert was used. Each expression vector was integrated in the *P. pastoris* genome under the control of the AOX1 promoter, and three independent clones of each GPCR construct and a negative vector-only control were analyzed. Yeast cultures were incubated at 22°C at 250 rpm for 15 h after induction, harvested by centrifuging the cultures at 1,500 × *g* at 4°C for 10 min, and the cells were resuspended in 5 ml TBS containing 1% protease inhibitor mixture (Sigma-Aldrich) and stored at −80°C. Details of membrane preparations are given in the *SI Text*.

Immunoblot Analysis. For immunoblot analysis, yeast membranes (≈20 μg of total membrane protein) were diluted with TBS to a concentration of 4 μg/μl and an equivalent volume of 2× SDS loading buffer was added. The samples were incubated at 42°C for 30 min in the presence of 10 mM DTT before separating the proteins by SDS/PAGE (4–12% Bis-Tris gels; Invitrogen). Proteins were transferred to Immobilon-P transfer membranes (Millipore), and membranes were blocked in TBST (TBS with 0.5% Tween-20) with 5% milk powder for 1 h at room temperature. The C-terminal biotinylation was detected with a streptavidin-alkaline phosphatase conjugate (Roche Diagnostic GmbH). While the expected size of the GPCR is ≈55 kDa, results from N-terminal sequencing and from in-gel digestions followed by mass spectrometry of the *P. pastoris* construct showed that the band at ≈43 kDa indeed corresponds to the full-length GPCR, correctly processed at the N terminus (data not shown).

Receptor Solubilization and Purification. This was performed essentially as published by Grisshammer and Tucker (30). The detailed protocol is provided in the *SI Text*.

Thermal Stability. Receptors were expressed as fusion proteins (MBP-GPCR-TrxA-His_{10}) in *E. coli* and purified by IMAC and size exclusion chromatography (Superdex 200). Thermal stability was assayed in buffer STAB30 [50 mM Tris•HCl, pH 7.4, 30% glycerol, 200 mM NaCl, 1 mM EDTA, 0.05% dodecyl-β-D-maltopyranoside (DDM), 0.5% (wt/vol) 3-[(3-cholamidopropyl)-dimethylammonio]-1-propane sulfonate (CHAPS), 0.1% (wt/vol); cholesteryl hemisuccinate (CHS)]. Samples were incubated at 45°C for the indicated period (up to 175 min) and were then put on ice. Radioligand binding assays were performed as described in the *SI Text*.

ACKNOWLEDGMENTS. The authors are grateful to Eva Niederer for invaluable assistance with flow cytometry, to Dr. Reinhard Grisshammer (National Institutes of Health) for helpful discussions and sharing vectors, and to Dr. Randy Schekman (University of California Berkeley) for antiprepro alpha factor antibodies. This work was supported by a National Institutes of Health postdoctoral fellowship (to C.A.S.), a predoctoral fellowship from the Forschungskredit of the University of Zurich (to I.D.), and by grants from the Transregio Program and from the Swiss National Science Foundation (NCCR in Structural Biology) (to A.P.).

1. Lundström K (2005) Structural genomics of GPCRs. *Trends Biotechnol* 23:103–108.
2. Berman HM, *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242.
3. Palczewski K, *et al.* (2000) Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* 289:739–745.
4. Okada T, *et al.* (2000) X-ray diffraction analysis of three-dimensional crystals of bovine rhodopsin obtained from mixed micelles. *J Struct Biol* 130:73–80.
5. Rasmussen SG, *et al.* (2007) Crystal structure of the human β_2 adrenergic G-protein-coupled receptor. *Nature* 450:383–387.
6. Cherezov V, *et al.* (2007) High-resolution crystal structure of an engineered human β_2-adrenergic G protein-coupled receptor. *Science* 318:1258–1265.
7. Sarramegna V, Talmont F, Demange P, Milon A (2003) Heterologous expression of G-protein-coupled receptors: Comparison of expression systems from the standpoint of large-scale production and purification. *Cell Mol Life Sci* 60:1529–1546.
8. Chen G, *et al.* (2001) Isolation of high-affinity ligand-binding proteins by periplasmic expression with cytometric screening (PECS). *Nat Biotechnol* 19:537–542.
9. Harvey BR, *et al.* (2004) Anchored periplasmic expression, a versatile technology for the isolation of high-affinity antibodies from *Escherichia coli*-expressed libraries. *Proc Natl Acad Sci USA* 101:9193–9198.
10. Grisshammer R, Duckworth R, Henderson R (1993) Expression of a rat neurotensin receptor in *Escherichia coli. Biochem J* 295:571–576.
11. Tucker J, Grisshammer R (1996) Purification of a rat neurotensin receptor expressed in *Escherichia coli. Biochem J* 317:891–899.
12. Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat Biotechnol* 16:258–261.
13. Gully D, *et al.* (1993) Biochemical and pharmacological profile of a potent and selective nonpeptide antagonist of the neurotensin receptor. *Proc Natl Acad Sci USA* 90:65–69.
14. Grisshammer R, Hermans E (2001) Functional coupling with $G\alpha_q$ and $G\alpha_{i1}$ protein subunits promotes high-affinity agonist binding to the neurotensin receptor NTS-1 expressed in *Escherichia coli. FEBS Lett* 493:101–105.
15. Rovati GE, Capra V, Neubig RR (2007) The highly conserved DRY motif of class A G protein-coupled receptors: Beyond the ground state. *Mol Pharmacol* 71:959–964.
16. Julius D, Schekman R, Thorner J (1984) Glycosylation and processing of prepro-α-factor through the yeast secretory pathway. *Cell* 36:309–318.
17. Shusta EV, Kieke MC, Parke E, Kranz DM, Wittrup KD (1999) Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency. *J Mol Biol* 292:949–956.
18. Kimchi-Sarfaty C, *et al.* (2007) A ''silent'' polymorphism in the MDR1 gene changes substrate specificity. *Science* 315:525–528.
19. Lin-Chao S, Chen WT, Wong TT (1992) High copy number of the pUC plasmid results from a Rom/Rop-suppressible point mutation in RNA II. *Mol Microbiol* 6:3385–3393.
20. Labbe-Jullie C, *et al.* (1998) Mutagenesis and modeling of the neurotensin receptor NTR1: Identification of residues that are critical for binding SR 48692, a nonpeptide neurotensin antagonist. *J Biol Chem* 273:16351–16357.
21. Barroso S, Richard F, Nicolas-Etheve D, Kitabgi P, Labbe-Jullie C (2002) Constitutive activation of the neurotensin receptor 1 by mutation of Phe^{358} in helix seven. *Br J Pharmacol* 135:997–1002.
22. Ault AD, Broach JR (2006) Creation of GPCR-based chemical sensors by directed evolution in yeast. *Protein Eng Des Sel* 19:1–8.
23. Sommers CM, *et al.* (2000) A limited spectrum of mutations causes constitutive activation of the yeast α-factor receptor. *Biochemistry* 39:6898–6909.
24. Newstead S, Kim H, von Heijne G, Iwata S, Drew D (2007) High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in *Saccharomyces cerevisiae. Proc Natl Acad Sci USA* 104:13936–13941.
25. Li B, *et al.* (2007) Rapid identification of functionally critical amino acids in a G protein-coupled receptor. *Nat Methods* 4:169–174.
26. Scarselli M, Li B, Kim SK, Wess J (2007) Multiple residues in the second extracellular loop are critical for M3 muscarinic acetylcholine receptor activation. *J Biol Chem* 282:7385–7396.
27. Sommers CM, Dumont ME (1997) Genetic interactions among the transmembrane segments of the G protein coupled receptor encoded by the yeast STE2 gene. *J Mol Biol* 266:559–575.
28. Molina DM, *et al.* (2008) Engineering membrane protein overproduction in *Escherichia coli. Protein Sci* 17:673–680.
29. Serrano-Vega MJ, Magnani F, Shibata Y, Tate CG (2008) Conformational thermostabilization of the β1-adrenergic receptor in a detergent-resistant form. *Proc Natl Acad Sci USA* 105:877–882.
30. Grisshammer R, Tucker J (1997) Quantitative evaluation of neurotensin receptor purification by immobilized metal affinity chromatography. *Protein Expr Purif* 11:53–60.
31. DuBridge RB, *et al.* (1987) Analysis of mutation in human cells by using an Epstein–Barr virus shuttle system. *Mol Cell Biol* 7:379–387.
32. Ott D, Frischknecht R, Plückthun A (2004) Construction and characterization of a kappa opioid receptor devoid of all free cysteines. *Protein Eng Des Sel* 17:37–48.
33. Vermeiren C, *et al.* (2006) Loss of metabotropic glutamate receptor-mediated regulation of glutamate transport in chemically activated astrocytes in a rat model of amyotrophic lateral sclerosis. *J Neurochem* 96:719–731.
34. Talmont F, Sidobre S, Demange P, Milon A, Emorine LJ (1996) Expression and pharmacological characterization of the human μ-opioid receptor in the methylotrophic yeast *Pichia pastoris. FEBS Lett* 394:268–272.

4.2 Supporting information

Sarkar *et al.* 10.1073/pnas.0803103105

SI Text

Synthesis of Fluorescent Neurotensin and Construction of Neurotensin Receptor Library. A fluorescent neurotensin analog was prepared by coupling BODIPY FL-X, SE (Invitrogen) to the minimal bioactive peptide sequence NT(8–13) (AnaSpec) according to the manufacturer's instructions. The conjugated product was isolated by reverse-phase HPLC and purity was confirmed by mass spectrometry.

The initial NTR1 library was generated by epPCR using regular dNTPs (200 μM) and the nucleotide analogs dPTP and 8-oxo-dGTP (1–15 μM in different libraries, which were then combined, corresponding to 3–13 amino acid mutations, on average) (1). Library sequencing after the first four rounds of FACS revealed only 2 to 3 amino acid mutations per sequence; therefore, additional epPCR steps were only performed using 1 μM analogs. DNA shuffling in the last randomization step was performed using the StEP (2). The library complexity was $\approx 10^6$ to 10^7 initial transformants after each randomization step, and the library size screened during each round of FACS consisted of $\approx 10^7$ to 10^8 cells.

Optimization of Binding Buffer for Selections. Previous work by Georgiou and colleagues (3) suggested that the outer membrane of *E. coli* could be made permeable to relatively small ligands (<10 kDa) while retaining viability by using concentrated salt solutions. Because our NTR1 variants are functionally expressed in the inner membrane of *E. coli*, we tested the efficacy of several permeabilization buffers allowing binding of [^{3}H]-NT added to the extracellular medium (Fig. S1).

From a simple thermodynamic model that accounts for partitioning of ligand across a selective membrane, it can be shown from the above data that the ligand concentration in the periplasm is roughly one fifteenth that in the extracellular medium when 5× Tris-KCl is used (data not shown). Correspondingly, we predicted that expressed receptors would be essentially saturated in this buffer when 50 nM [^{3}H]-NT was used.

At this higher concentration of ligand, complete receptor saturation is achieved with the standard buffer and the extent of specific receptor binding is almost fully matched by the 5× Tris-KCl buffer (Fig. S2). Notably, Na^+ is a known inhibitor of NT in binding NTR1, so it is not surprising that the extent of receptor binding is reduced in the buffer containing high concentrations of Na^+. Additionally, divalent cations do not appear to maximize membrane permeability. Therefore, a buffer consisting of Tris and KCl was selected for further refinement. The influence of buffer concentration and composition on receptor binding (Fig. S3), enrichment ratios in FACS (Fig. S4), and viability of recovered cells after FACS (see Fig. S4) was determined. It is clear from Fig. S3 that Tris is critical for maximizing the signal, although more concentrated buffer solutions are not necessary. Then, cells incubated in the standard assay buffer, 5× Tris-KCl, 1× Tris, and 1× Tris-KCl were subject to FACS. To set the gating window, cells were incubated with saturating amounts of BODIPY-NT(8–13) but also with excess unlabeled NT(8–13), and the left boundary of the gating window was positioned just to the right of the cell populations for the standard buffer, 1× Tris, and 1× Tris-KCl (see Fig. S4). The 5× Tris-KCl buffer was not given further consideration, as it was evident from the scatter plots that the forward scatter (and thus the cell morphology) was significantly altered by the buffer. The signal-to-noise ratio in the gating window for the 1× Tris-KCl buffer was ≈900 and this buffer gave the best cell viability; therefore, it was chosen for all future experiments.

Flowchart for Selections. The procedure for performing selections for expression level and selectivity was as shown in Fig. S5.

Single-Clone Analysis After Selections. After the selections for expression and selectivity, individual clones were isolated for sequencing and expression. As outlined in Fig. S5, after the third round of epPCR and selection, enriched sequences were re-randomized either by a fourth round of epPCR or by shuffling through StEP and then screened for either maximal expression or ligand selectivity. Then, individual clones were chosen from each library in each of the selections [48 epPCR clones and 48 StEP clones in maximal expression (Fig. S6), and 48 of each in altered binding selectivity (Fig. S7)] and all 192 were compared to NTR1 for expression level by performing radioligand binding assays as outlined in *Methods*.

The properties of the best expressing clones from this analysis, D03 for maximal expression and G10 for altered ligand selectivity (both from 4EP03, not StEP), were also checked by flow cytometry (Figs. S8 and S9, respectively).

Amino Acid Mutations in D03 and G10. The positions of the evolved mutations in D03 are shown in the snake plot in Fig. S10. The location of the key selectivity mutation in G10 (F358S), which does not exist in D03, is also highlighted on this template.

Radioligand Binding Assays. Whole-cell binding assays with *E. coli* (20 μl culture) were performed with 380 μl of assay buffer [50 mM Tris·HCl, pH 7.4, 1 mM EDTA, 40 μg/ml bacitracin, 0.1% (wt/vol) BSA] containing a fixed concentration (10 nM) or dilution series (0.04–20 nM) of radioactive agonist [^{3}H]-NT (Perkin-Elmer), essentially as described (4). The expression culture was cooled to 4°C for at least 2 h before addition of assay buffer; thereafter, the mixed sample (culture plus assay buffer) was incubated at 4°C for an additional 2 h, centrifuged, washed and measured by scintillation counting. Nonspecific binding was determined in the presence of 5 μM unlabeled NT. Competition with antagonist SR 48692 (Sanofi Aventis) was determined at a concentration of 5 μM.

To measure binding to mammalian cells, cell pellets were resuspended in 500 μl of assay buffer with a fixed (2 nM) or increasing concentration of radioactive agonist [^{3}H]-NT (1–20 nM). Nonspecific binding was determined in the presence of 1 μM unlabeled NT. The cell suspension (250 μl) was incubated at room temperature for 20 min and the assay was terminated with 1 ml of ice-cold assay buffer and the free ligand was separated from bound ligand using 96-well filtration plates with glass-fiber filters (MAFBNOB, Millipore) pretreated with 0.5% (vol/vol) polyethylenimine and a vacuum filtration manifold (Millipore). OptiPhase HiSafe 2 scintillation mixture (Perkin-Elmer) was added directly to the wells (100 μl/well). The plate was shaken overnight and radioactivity was counted in a liquid scintillation counter (1450 Microbeta Plus, Perkin-Elmer).

Experiments with membranes from *P. pastoris* were performed similarly, but with a few modifications. The binding reaction took place in a thermomixer for 1 h at 30°C and 1,400 rpm in 250 μl binding buffer containing ≈70 μg of total membrane protein fractions (diluted with Tris-buffered saline (TBS) to ≈14 μg/μl) and 15 nM [^{3}H]-NT. Nonspecific binding was determined in the presence of 40 μM unlabeled NT. After

binding, 200 μl of membrane suspension was transferred to a 96-well filtration plate on the vacuum manifold. The filters were washed five times with 200 μl of cold assay buffer.

For solubilized receptor, an aliquot (1.5 μl) was incubated with 150 μl assay buffer containing 50 nM [^{3}H]-NT and detergents (0.05% (wt/vol) DDM, 0.5% (wt/vol) CHAPS, 0.1% (wt/vol) CHS; Anatrace) for 1.5 h at 4°C. Separation of receptor-ligand complex from free ligand was achieved with gel-filtration spin columns [Bio-Spin 30 (Tris buffer) column, Bio-Rad] preequilibrated with detergents. Radioactivity was measured by liquid scintillation counting.

***P. pastoris* Membrane Preparation.** For membrane preparations, cells were thawed on ice, and 5 ml of acid-washed glass beads (425–600 microns; Sigma-Aldrich) were added to the cell suspensions. Cell disruption was performed in three cycles (5 min at maximal speed, 5 min on ice to cool) in a TissueLyser (Qiagen). The membrane suspensions were transferred to 15-ml tubes and centrifuged at 4,000 × *g* for 15 min at 4°C to remove cell debris. The membrane suspensions were transferred to new tubes, and centrifuged at 150,000 × *g* for 30 min at 4°C. The membrane pellet was resuspended in 500 μl of ice-cold TBS containing 1% protease inhibitor mixture. The protein concentrations of the membrane preparations were determined by a Quant-iT Protein Assay Kit (Invitrogen) and the membrane suspensions were stored at −20°C.

Receptor Solubilization and Purification. Frozen cell pellets (−80°C) from 1 liter of expression culture (4–5 g) were thawed and resuspended in 10 ml 2× resuspension buffer (100 mM Tris·HCl, pH 7.4, 60% glycerol, 400 mM NaCl) with a disperser (IKA yellow line DI 18, 1 min, level 1). The following steps were carried out at 0–4°C. Protease inhibitors (mini complete protease inhibitors, EDTA-free, 1 tablet, Roche Applied Science), $MgCl_2$ (10 mM), DNase I (20 μg/ml, Roche Applied Science), and detergents [1.5% DDM; 0.5% CHAPS; 0.1% CHS] were added with gentle stirring. The volume was adjusted to 20 ml with ultrahigh purity water, sonicated for 3 min (Branson Sonifier 250, level 4, 30% duty cycle, ½-inch flat tip), and incubated for 2 h with gentle mixing. Cell debris was removed by centrifugation (20 min, 50,000 × *g*) followed by ultracentrifugation (45 min, 190,000 × *g*). Imidazole was then added to the clarified lysate to a final concentration of 50 mM, pH 7.8–7.9, before IMAC. Solubilized protein was passed at a flow rate of 0.3 ml/min over a 2-ml Superflow Ni^{2+}-NTA column (Qiagen) equilibrated with buffer NiA (50 mM Tris·HCl, pH 7.4, 30% glycerol, 200 mM NaCl, 50 mM imidazole, 0.1% DDM, 0.5% CHAPS, 0.1% CHS). The resin was washed with 15 column volumes of buffer NiA at a flow rate of 0.5 ml/min and the protein was eluted (0.2 ml/min) with buffer NiA containing 300 mM imidazole. For further purification by size exclusion chromatography (SEC), the collected peak fractions were pooled (4–5 ml) and concentrated to 250 μl (Amicon Ultra-4 50K, Millipore). The concentrated eluate was passed through a Superdex 200 column (GE Healthcare) equilibrated with buffer SEC15 (50 mM Tris·HCl, pH 7.4, 15% glycerol, 1 M NaCl, 1 mM EDTA, 0.05% DDM, 0.5% CHAPS, 0.1% CHS) at a flow rate of 0.3 ml/min. All chromatography steps were performed on an Äkta Prime system (GE Healthcare). Radioligand binding assays with solubilized receptor were performed as described above.

All variants tested bind quantitatively to a NT-affinity column (data not shown) (4), but D03 evolved to be refractive to competition with Na^+ and could therefore not be eluted from this column with high salt. That binding to the affinity column (monomeric avidin saturated with biotinylated NT) is indeed because of productive binding of ligand was shown by specifically eluting bound receptor with an excess of free biotin (data not shown). Consequently, affinity purification was omitted for comparative purposes, as elution with Na^+ was not possible for all variants.

$[Ca^{2+}]_i$ Signaling Patterns Mediated by NT. Common features of NT signaling that have been recognized are activation of phospholipase C (PLC), production of inositol-1,4,5-trisphosphate, and mobilization of intracellular calcium ($[Ca^{2+}]_i$), suggesting that this receptor is coupled to $G_{q/11}$ protein (5, 6). Activation of PLC-coupled GPCRs can trigger a variety of Ca^{2+} signaling patterns (7), and the nature of the resulting signal depends on factors such as cell type, extracellular concentration of Ca^{2+}, and choice of agonist and its concentration. At the lowest agonist concentrations at which any signaling is observed in HEK293T cells expressing NTR1 or its mutants, application of NT predominantly induced sinusoidal $[Ca^{2+}]_i$ oscillations. Oscillations in $[Ca^{2+}]_i$ have been observed during signal transduction for a wide variety of receptors (8, 9); these fluctuations may enable frequency-dependent cellular responses as opposed to amplitude-dependent responses, which would require sustained exposure to elevated $[Ca^{2+}]_i$. At higher concentrations, a strong but transient response (no oscillations) is observed. At supersaturating concentrations of agonist, maximal release of Ca^{2+} is achieved. This is then followed by store-depletion gated Ca^{2+} influx, resulting in typical plateau-shaped $[Ca^{2+}]_i$ response curves (10).

The increase in $[Ca^{2+}]_i$ upon agonist binding was measured in single-cell experiments, as described in *Methods*. Upon stimulation, three different types of response were observed at the single-cell level, even at a given concentration of agonist (Fig. S11).

The three responses are characterized as plateau, transient, and oscillatory. A plateau response (bold green trace in Fig. S11) is defined as one in which the fluorescence ratio does not decrease to the baseline value but remains at a value at least 30% of the peak maximum 90 seconds after treatment with agonist has been stopped. It has been suggested that this type of response is caused mainly by the influx of Ca^{2+} from the extracellular space (11). The plateau response occurs primarily at very high NT concentrations. At intermediate agonist concentrations, only minimal depletion of Ca^{2+} stores is thought to occur, and therefore a plateau response is less common. A transient response (bold blue trace in Fig. S11) is defined as one in which the fluorescence ratio increases shortly after receptor activation and returns approximately to the baseline value within 60 s. Finally, an oscillatory response is defined as one in which multiple transient responses arise within a period of 180 to 200 s after agonist activation. Oscillations in $[Ca^{2+}]_i$ occur under a variety of conditions and play functional roles in signal transduction and cell regulation (12). They typically arise from calcium-induced calcium release mechanisms, in which an increase in $[Ca^{2+}]_i$ leads to increased Ca^{2+} influx from intracellular Ca^{2+} stores (13). In this experimental system, the observed oscillations differed in both frequency and intensity. Most oscillating responses occurred at low (0.1 nM or less) neurotensin concentrations and disappeared almost entirely at high agonist concentrations (10 nM or more).

While plateau, transient, and oscillatory responses all exist at a given concentration of agonist, the likelihood of observing a particular type of response changes with the type of receptor and the agonist concentration (see main text and Fig. 3).

The selection of NTR1 variants was performed on N-terminally truncated receptor (amino acids 43–424). Therefore, receptor expression and signaling in mammalian cells were also performed with truncated versions. However, to show that receptor truncation does not alter the intrinsic signaling patterns, full-length NTR1 (NTR1fl) and D03 (D03fl) were constructed using a synthetic gene for the N-terminal domain. No major difference between the truncated and full-length forms was

detected. Oscillations are most dominant at low agonist concentrations (0.01 nM), and larger fractions of transient and plateau responses are observed as the agonist concentration is increased (see Fig. 3 and Fig. S12).

Expression of NTR1 Variants in *E. coli* and *P. pastoris*. To compare total GPCR biosynthesis per bacterial cell, NTR1, D03, and D03-L167R were expressed in parallel *E. coli* cultures, which were then lysed and probed for total full-length receptor fusion by whole-cell Western blotting (Fig. S13). The order-of-magnitude increases in D03 and D03-L167R expression levels seen in this blot (a measure of all full-length receptor molecules per cell) are in good agreement with the respective 8.7- and 6.6-fold increases observed by radioligand binding assay (a measure of folded, functional receptor in the inner bacterial membrane; see Table 1 in main text).

The enhanced expression of D03 that is observed in *P. pastoris* could result either from an intrinsic increase in protein biosynthesis of this variant or from more efficient folding after translocation and processing. To address this, we detected the precursor form of NTR1, D03, and D03-L167R expressed in *P. pastoris* by Western blot (Fig. S14).

Because this precursor form is detected at a similar intensity for all samples (except the "ctrl"), it is reasonable to conclude that the order-of-magnitude increase in expression for D03 in *P. pastoris* arises primarily from more efficient folding and not from enhanced biosynthesis.

Influence of Point Mutation in Expression Vector Backbone. During selection, a mutation close to the origin of replication in the expression vector (Fig. S15). To determine what effect, if any, this mutation [C7054G; position 1 corresponding to the first G nucleotide of the *lacI*^{q} gene of the vector pRG/III-hs-MBP (14)] might have on plasmid copy number and receptor expression, NTR1 and D03 were cloned into the original vector (denoted pRG_{WT}) and the mutant vector (denoted pRG_{C7054G}) (Fig. S16).

Based on Fig. S16, it is clear that the C7054G point mutation increases plasmid copy number. However, this increase in DNA content does not correlate to any increase in functional NTR1 protein expression, suggesting that other, more significant bottlenecks exist along the biosynthetic pathway. Conversely, the corresponding increase in D03 DNA does result in only ≈25% more functional protein, implying that other bottlenecks along the biosynthetic pathway have been at least partially alleviated by the improved biophysical properties of the D03 sequence.

Gel Filtration of NTR1 and D03. Gel filtration was performed on detergent-solubilized, purified NTR1 and D03 to compare the two elution profiles. As shown in Fig. S17, the two traces are virtually identical with respect to the dominant peak (V_e = 11.9 ml).

To further confirm that the dominant protein peak does indeed contain functional receptor, radioligand-binding assays were performed (as described above) with sample fractions that were collected from this peak. The result for D03 is given in Fig. S8, which directly shows that the amount of functional receptor in each collected fraction closely mirrors the absorbance signal of the eluate at the corresponding elution volume.

Effect of Additive and Subtractive Point Mutations in NTR1 and D03, Respectively. To ascertain how isolated amino acid substitutions might affect the expression of NTR and D03, each mutation in D03 was singly introduced into NTR1 and, similarly, each mutation in D03 was reverted back to the WT residue. Using whole-cell binding assays with saturating amounts of radioligand (see above), the expression level of these 28 NTR1 variants in *E. coli* was determined (Fig. S19). As seen in this figure, the effect of each substitution on expression level is relatively small. Nevertheless, the sum of these multiple minor enhancements results in a variant, D03, with an order-of-magnitude increase in expression level; we refer to this phenomenon as a "staircase effect." This is in stark contrast to single amino acid substitutions that have a dramatic impact on phenotype, as in the case of the selectivity variant G10 (see main text); we refer to this phenomenon as an "elevator effect."

1. Zaccolo M, Gherardi E (1999) The effect of high-frequency random mutagenesis on in vitro protein evolution: A study on TEM-1 β-lactamase. *J Mol Biol* 285:775–783.
2. Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat Biotechnol* 16:258–261.
3. Chen G, *et al.* (2001) Isolation of high-affinity ligand-binding proteins by periplasmic expression with cytometric screening (PECS). *Nat Biotechnol* 19:537–542.
4. Tucker J, Grisshammer R (1996) Purification of a rat neurotensin receptor expressed in. *Escherichia coli Biochem J* 317:891–899.
5. Kitabgi P (2006) Functional domains of the subtype 1 neurotensin receptor (NTS1). *Peptides* 27:2461–2468.
6. Hermans E (2003) Biochemical and pharmacological control of the multiplicity of coupling at G-protein-coupled receptors. *Pharmacol Ther* 99:25–44.
7. Rovati GE, Capra V, Neubig RR (2007) The highly conserved DRY motif of class A G protein-coupled receptors: Beyond the ground state. *Mol Pharmacol* 71:959–964.
8. Hu Q, Deshpande S, Irani K, Ziegelstein RC (1999) [Ca^{2+}]$_i$ oscillation frequency regulates agonist-stimulated NF-κB transcriptional activity. *J Biol Chem* 274:33995–33998.
9. Berridge MJ, Lipp P, Bootman MD (2000) The versatility and universality of calcium signaling. *Nat Rev Mol Cell Biol* 1:11–21.
10. Liu X, Ambudkar IS (2001) Characteristics of a store-operated calcium-permeable channel: Sarcoendoplasmic reticulum calcium pump function controls channel gating. *J Biol Chem* 276:29891–29898.
11. Suh SH, *et al* (1999) Characterisation of explanted endothelial cells from mouse aorta: Electrophysiology and Ca^{2+} signalling. *Pflügers Arch* 438:612–620.
12. Berridge MJ, Bootman MD, Roderick HL (2003) Calcium signalling: Dynamics, homeostasis and remodeling. *Nat Rev Mol Cell Biol* 4:517–529.
13. Goldbeter A, Dupont G (1990) Allosteric regulation, cooperativity, and biochemical oscillations. *Biophys Chem* 37:341–353.
14. White JF, Trinh LB, Shiloach J, Grisshammer R (2004) Automated large-scale purification of a G protein-coupled receptor for neurotensin. *FEBS Lett* 564:289–293.

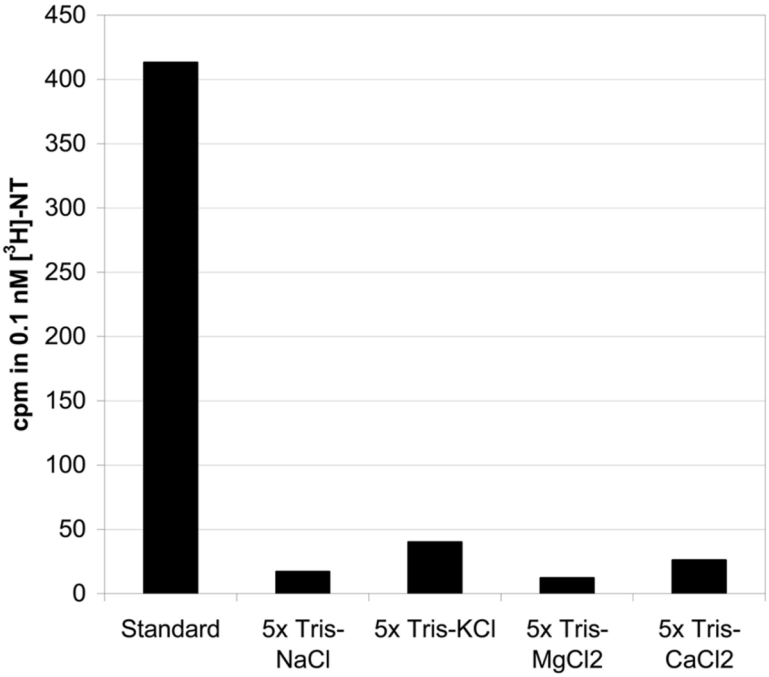


Figure S1: Specific binding of 0.1 nM [^{3}H]-NT to cells expressing NTR1 in the specific buffers. ‘‘Standard’’ is the normal binding assay buffer, defined in the Methods section, and contains 1 mM EDTA, which renders the outer membrane fully permeable. A 5x buffer formulation comprises 250 mM Tris, 750 mM chloride salt, pH 7.4. Note that the extracellular ligand concentration used here is approximately the K_d value.

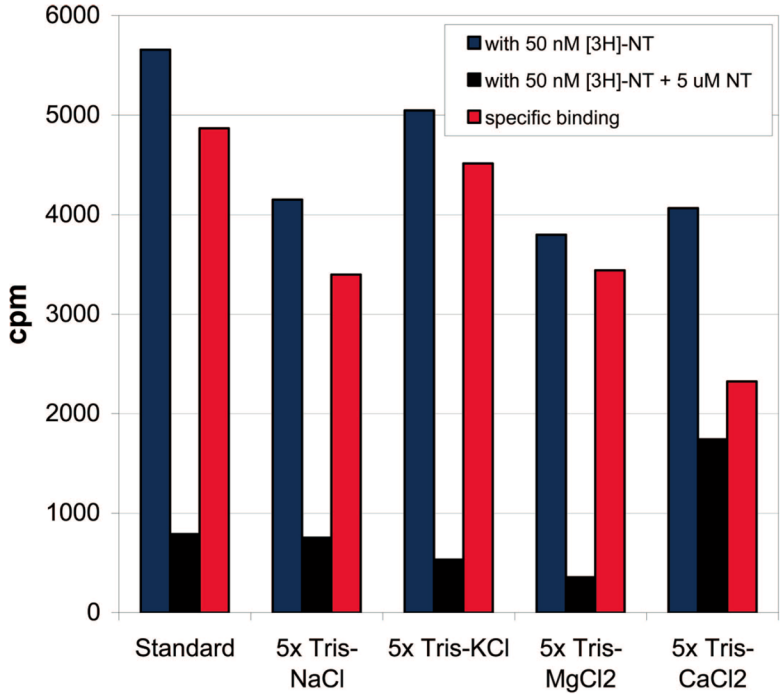


Figure S2: Total, nonspecific, and specific binding of 50 nM [^{3}H]-NT to cells expressing NTR1 in the specific buffers. ‘‘Standard’’ is the normal binding assay buffer, defined in the Methods section, and contains 1 mM EDTA, which renders the outer membrane fully permeable. A 5x buffer formulation comprises 250 mM Tris, 750 mM chloride salt, pH 7.4.

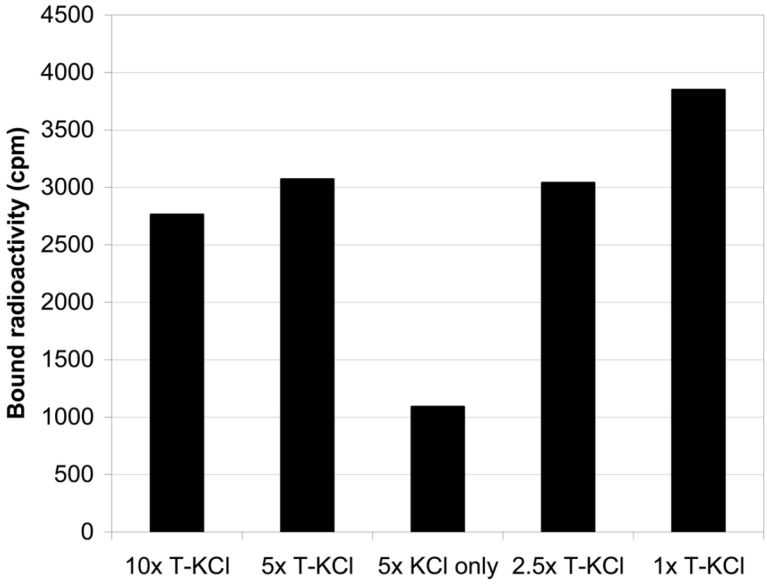


Figure S3: Influence of buffer concentration (and composition) on specific binding of 50 nM [^{3}H]-NT to cells expressing NTR1. Note that Tris (abbreviated ‘‘T’’) is critical for efficient permeabilization of the outer membrane, as evidenced by the sample with 5x KCl only.

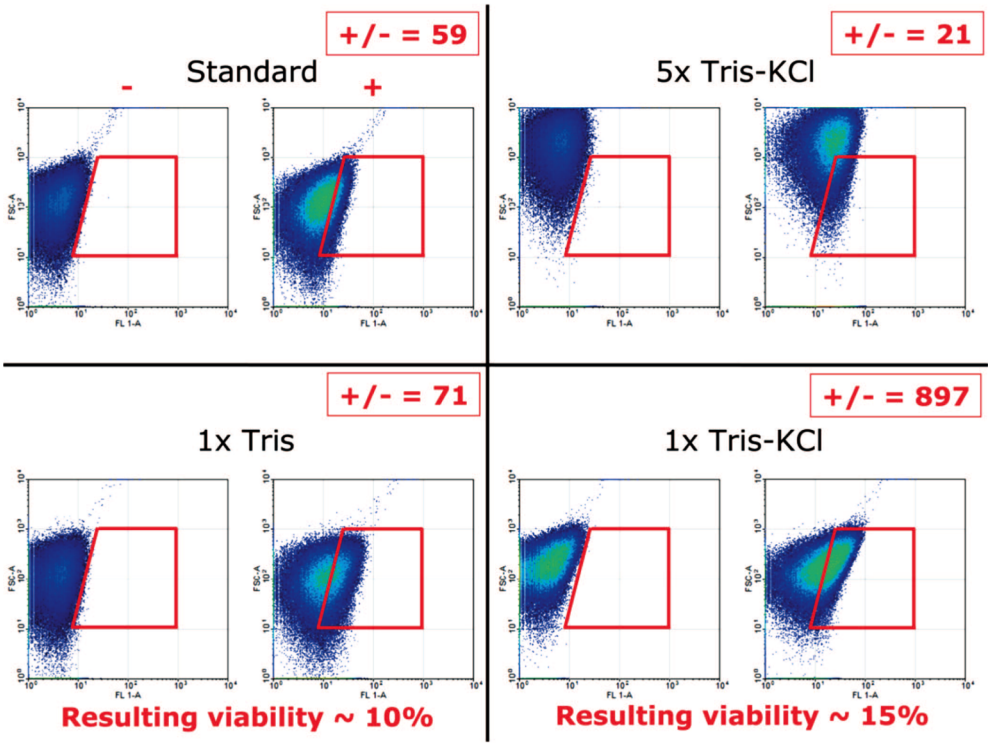


Figure S4: Influence of buffer concentration and composition on enrichment ratios in a specified gate in FACS. Scatter plots show green fluorescence on the x-axis and forward scatter on the y-axis. The same gate was used in all eight samples shown. In comparison to the other samples, the 5x Tris-KCl buffer appeared to alter cell morphology, as detected by shifted forward scatter. Cell viability was determined by dispensing an aliquot of recovered cells from FACS onto ampicillin-containing agar plates and comparing the number of viable colonies to the expected number of cells in that volume.

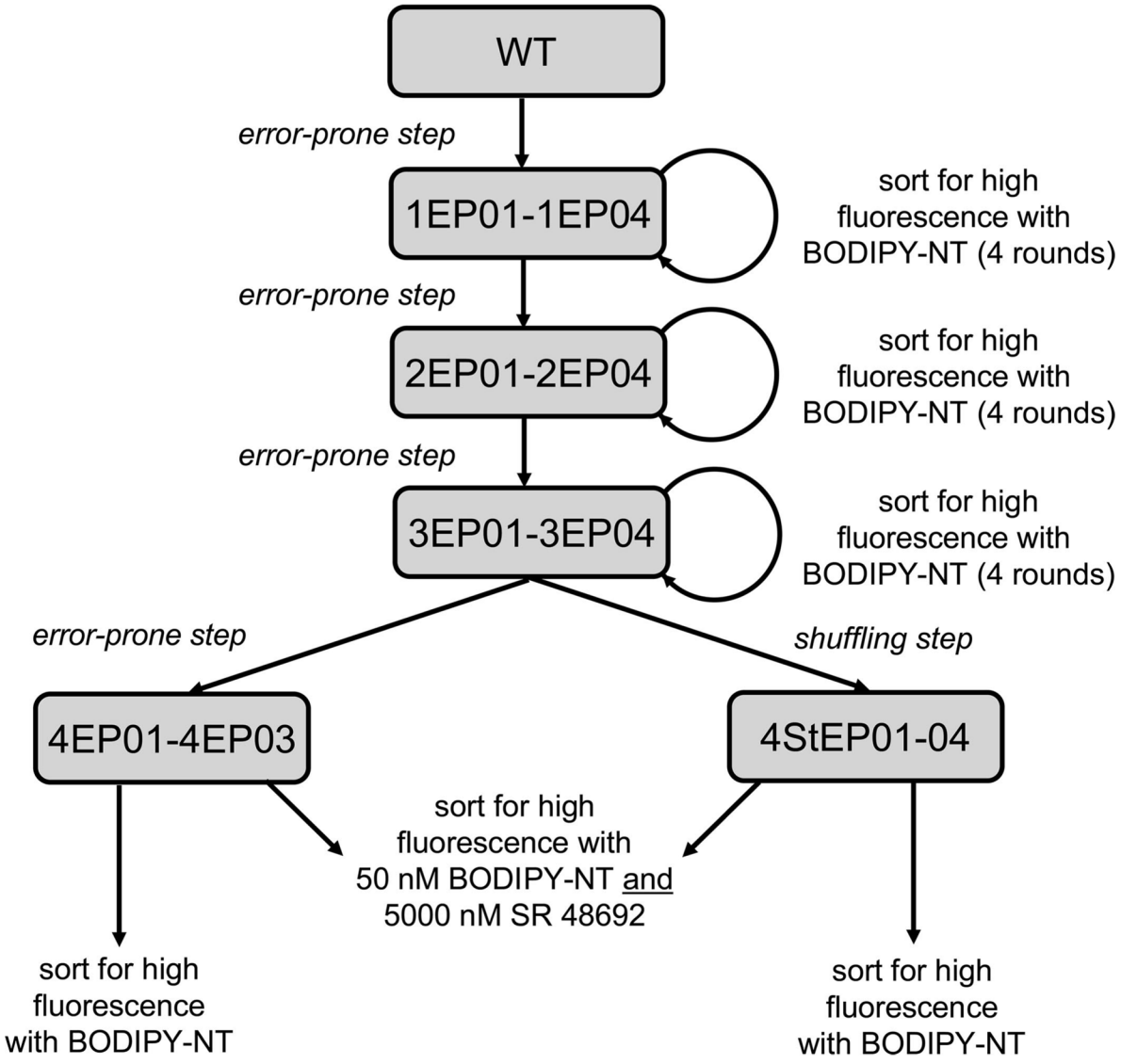


Figure S5: Flowchart for selections of NTR1 variants with increased expression level or altered ligand selectivity.

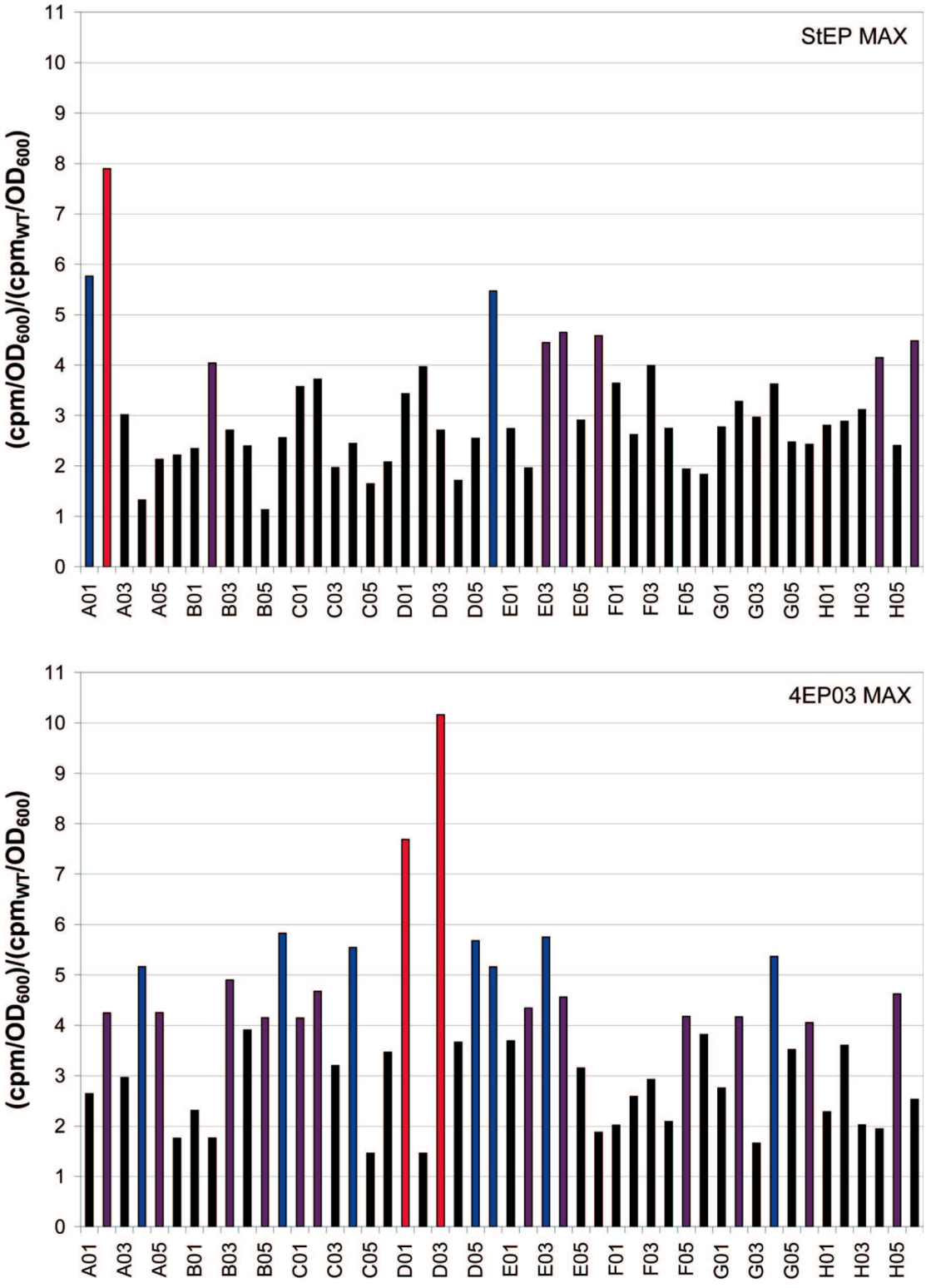


Figure S6: Radioligand binding analysis of 96 clones isolated from selections for maximum receptor expression level. The y-axis represents receptor expression level relative to wild-type NTR1 expression level. The 48 clones in the top plot were chosen from the StEP library; the 48 in the bottom from the epPCR library. Purple bars represent clones with expression levels at least fourfold that of NTR1; blue bars at least fivefold; and red bars at least sevenfold. D03 from 4EP03 MAX is the clone with the highest expression level in this experiment (10-fold that of NTR1).

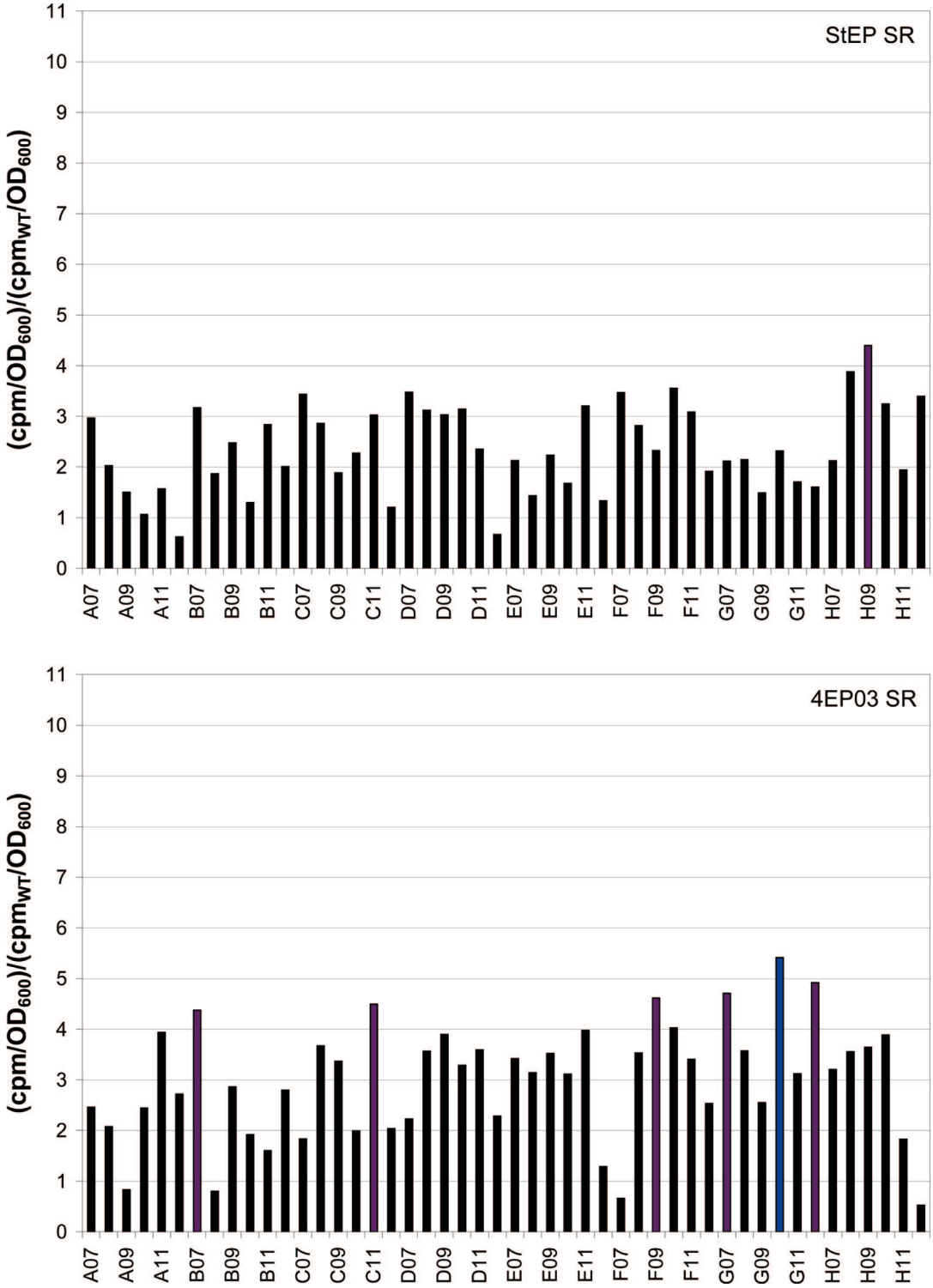


Figure S7: Radioligand binding analysis of 96 clones isolated from selections for altered ligand selectivity. The y-axis represents receptor expression level relative to WT NTR1 expression level. The 48 clones in the top plot were chosen from the StEP library; the 48 in the bottom from the epPCR library. Purple bars represent clones with expression levels at least fourfold that of NTR1 and blue bars at least fivefold. G10 from 4EP03 SR is the clone with the highest expression level in this experiment (fivefold that of NTR1).

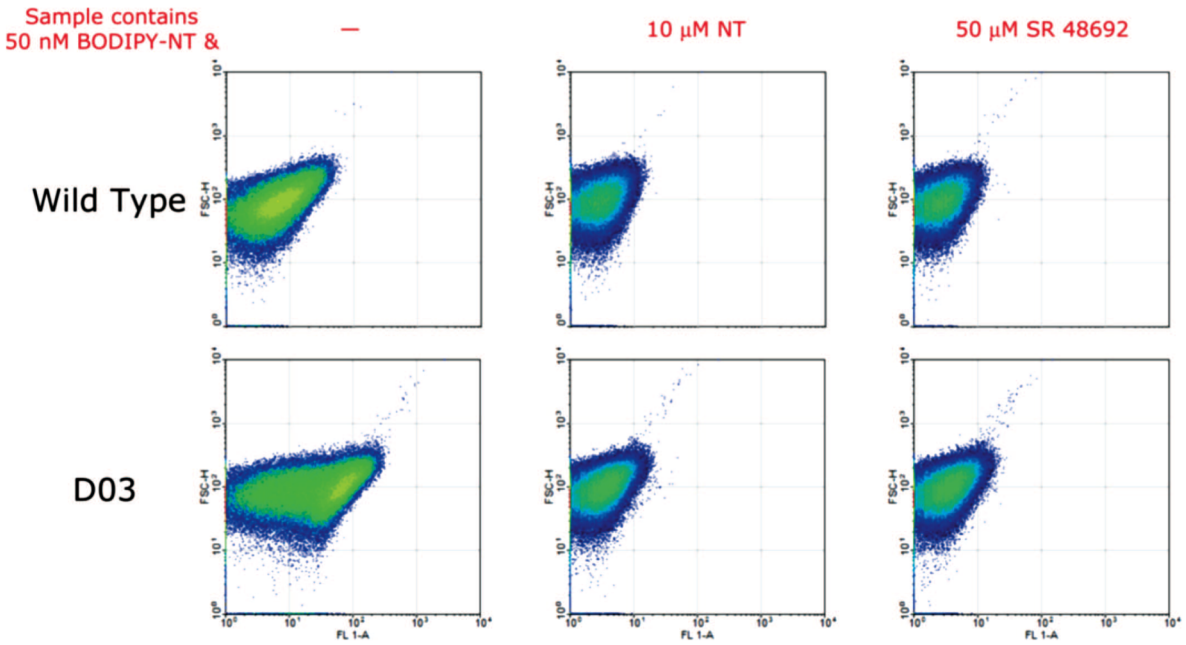


Figure S8: Flow cytometric analysis of D03 compared to WT NTR1. The mean fluorescence signal obtained for D03 in the presence of saturating BODIPY-NT alone is an order of magnitude greater than that of NTR1. However, this signal can be fully competed by addition of either excess NT or excess SR 48692.

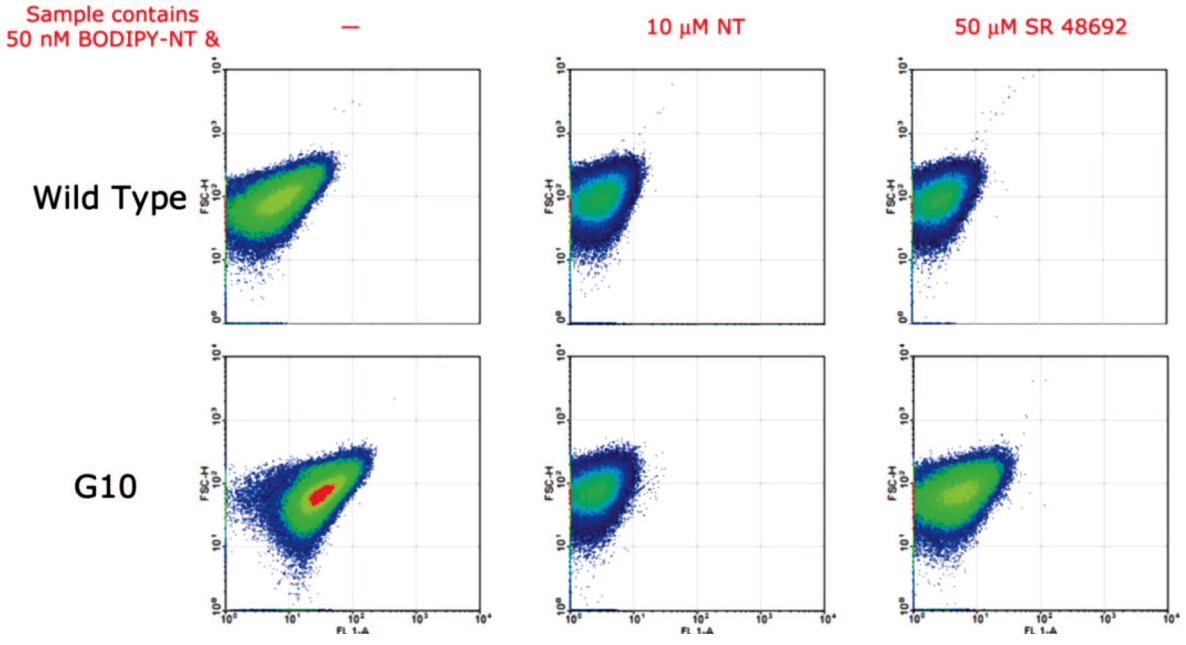
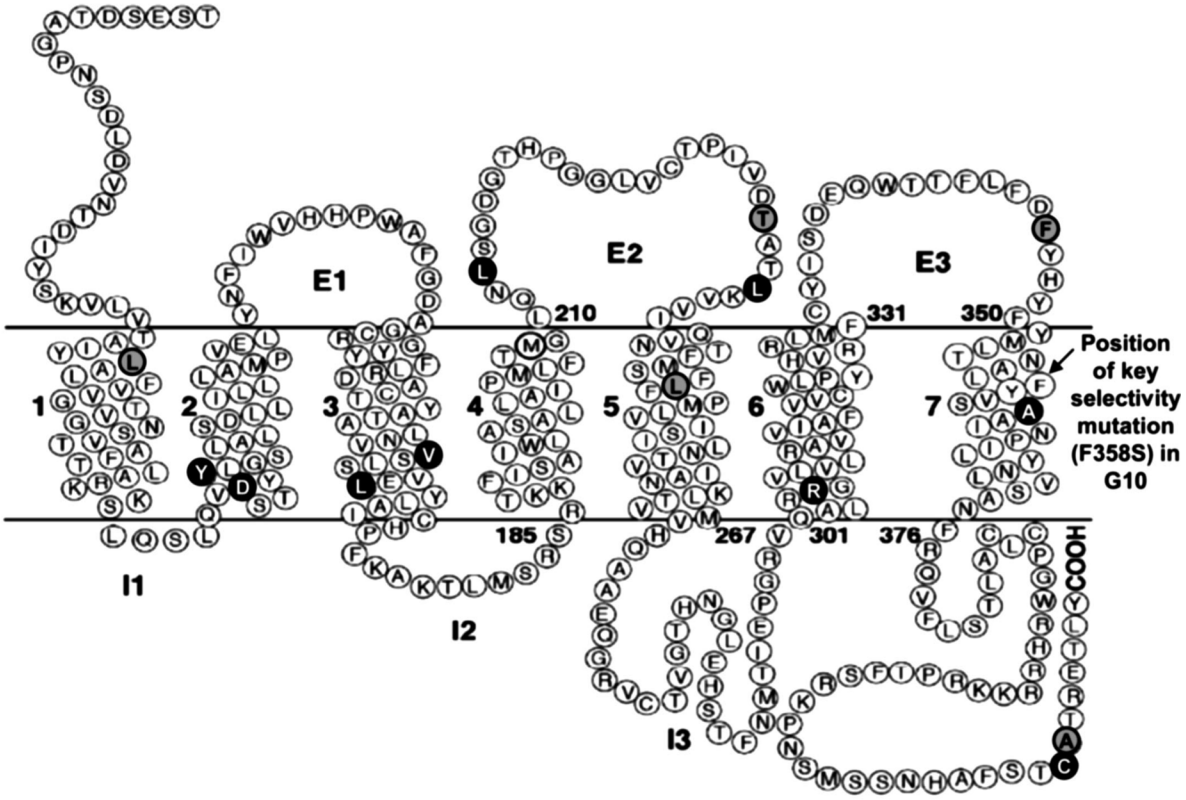


Figure S9: Flow cytometric analysis of G10 compared to WT NTR1. The mean fluorescence signal obtained for G10 in the presence of saturating BODIPY-NT alone is approximately fivefold that of NTR1. This fluorescence signal can be fully competed by addition of excess NT; however, addition of excess SR 48692 does not achieve full inhibition.



● Silent mutations

Leu72	(CTG)	→ Leu	(CTA^{rare})
Thr231	(ACA)	→ Thr	(ACT)
Leu247	(CTG)	→ Leu	(CTA^{rare})
Phe346	(TTC)	→ Phe	(TTT)
Ala418	(GCC)	→ Ala	(GCT)

● Non-silent mutations

His103	(CAT)	→ Asp	(GAT)
His105	(CAC)	→ Tyr	(TAC)
Ala161	(GCC)	→ Val	(GTC)
Arg167	(CGC)	→ Leu	(CTC)
Arg213	(CGC)	→ Leu	(CTC)
Val234	(GTC)	→ Leu	(CTC)
His305	(CAC)	→ Arg	(CGC)
Ser362	(TCC)	→ Ala	(GCC)
Ser417	(AGC)	→ Cys	(TGC)

Figure S10: Snake plot of NTR1 highlighting the silent (gray) and nonsilent (black) mutations in D03. The location of the key selectivity mutation in G10 (F358S) is also highlighted on this plot, although this mutation does not exist in D03.

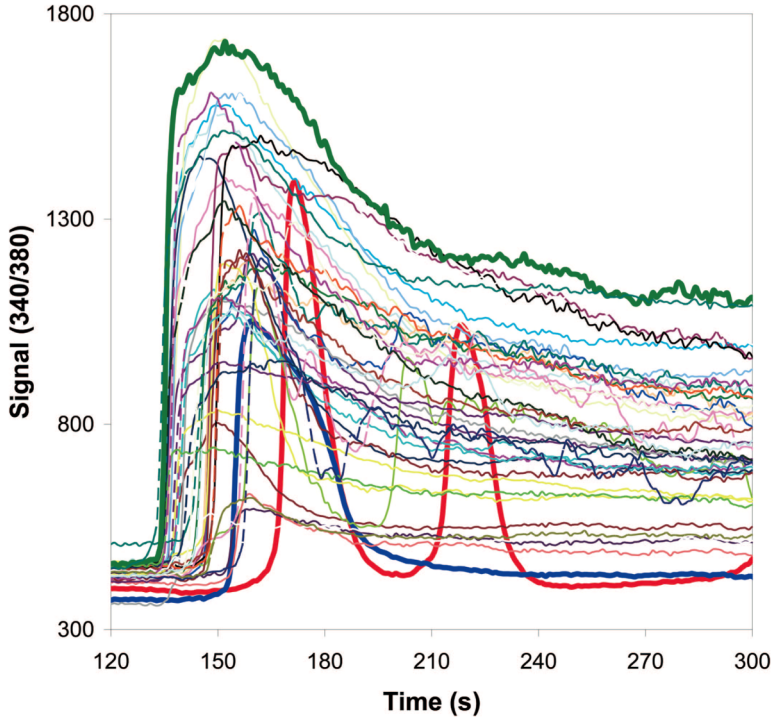


Figure S11: Variation in cytosolic calcium levels in individual HEK293T cells in response to 1 nM NT. Each trace represents a single cell and is characterized as plateau (bold green), transient (bold blue), or oscillatory (bold red).

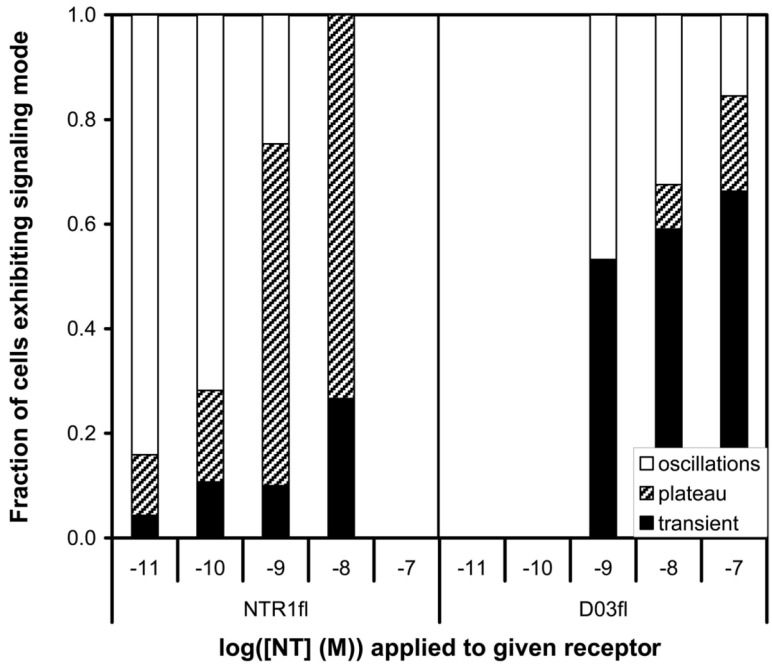


Figure S12: Ca^{2+} signaling in HEK293T cells expressing NTR1fl or D03fl (fl, full length, i.e., no N-terminal truncation). A comparison with **Fig. 3** in the main text suggests that the N-terminal truncation has minimal effect on the nature of signal transduction through these receptors. Oscillatory, plateau, and transient responses are illustrated in **Fig. S11** and defined in the SI Text.

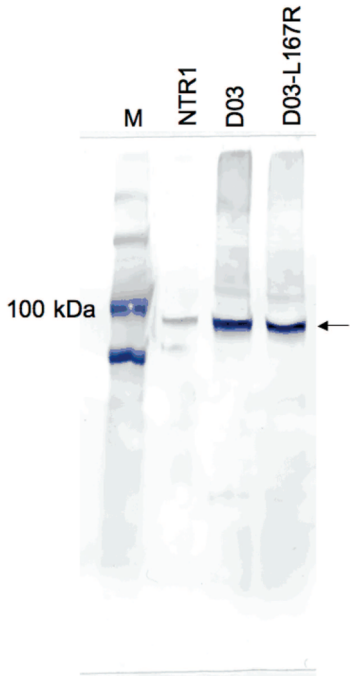


Figure S13: Whole-cell Western blot of NTR1, D03, and D03-L167R expressed in *E. coli*. Each lane was loaded with 5 µl of cell suspension (adjusted to OD_{600} = 25). The C-terminal His10 tag was detected with a primary mouse anti-His4 antibody and a secondary anti-mouse IgG alkaline phosphatase conjugate. The arrow indicates the recombinant fusion protein (MBP-GPCR-TrxA-His10) with the expected molecular mass of 100.2 kDa.

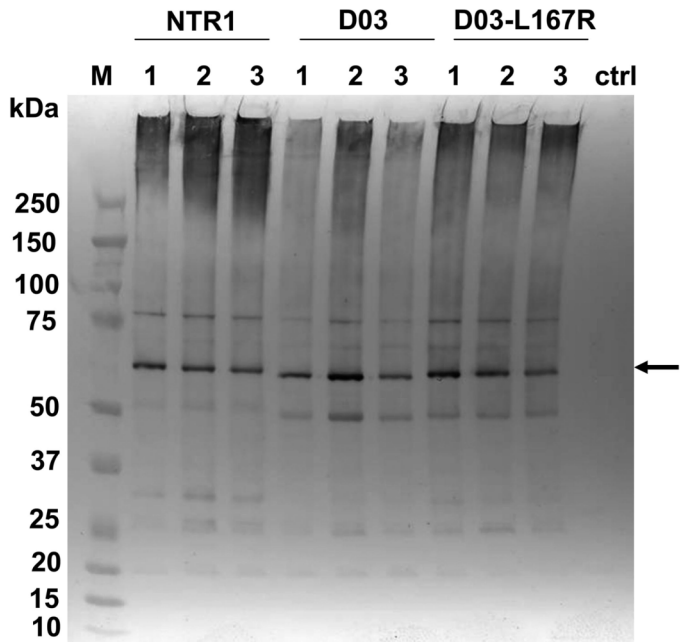


Figure S14: NTR1, D03, and D03-L167R were analyzed by Western blot. In each lane, 20 µg of total membrane protein was loaded. The GPCR precursor form was detected with anti-prepro-α-factor antiserum [primary; courtesy of Randy Schekman (University of California, Berkeley)] and alkaline phosphatase conjugated goat anti-rabbit-IgG (secondary). The arrow on the right of the blot denotes the size corresponding to this precursor form. The negative control (ctrl) lacks a GPCR insert.

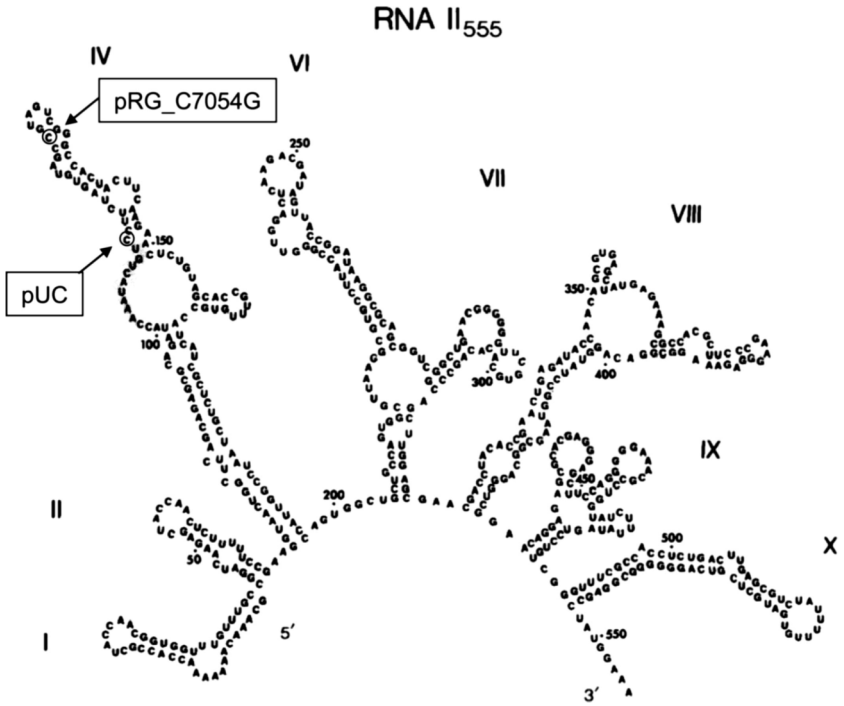


Figure S15: Position of the mutation C7054G in the secondary structure of RNA II555 of ColE1-harboring plasmid pRG. The structural model shows that C7054G is positioned near the mutation that confers high levels of plasmid replication to pUC derived plasmids and close to the region which is complementary to the 5' tail region of RNA I. [Reproduced with permission from ref. 1 (Copyright 1986, Elsevier).]

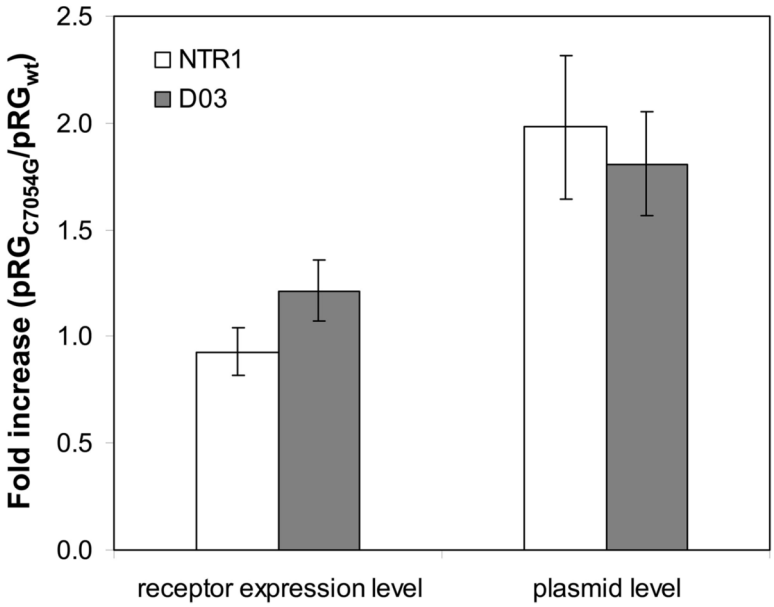


Figure S16: Effect of the mutation C7054G (located in the origin of replication of the plasmid pRG) on receptor expression and plasmid production levels in *E. coli*. Receptor expression increases for D03 but not for NTR1, when expression is driven from the mutated plasmid pRG_C7054G. C7054G leads to a twofold increase in cellular plasmid content independent of the receptor gene cloned into the vector. All ratios were calculated based on data that had been normalized for OD_{600}.

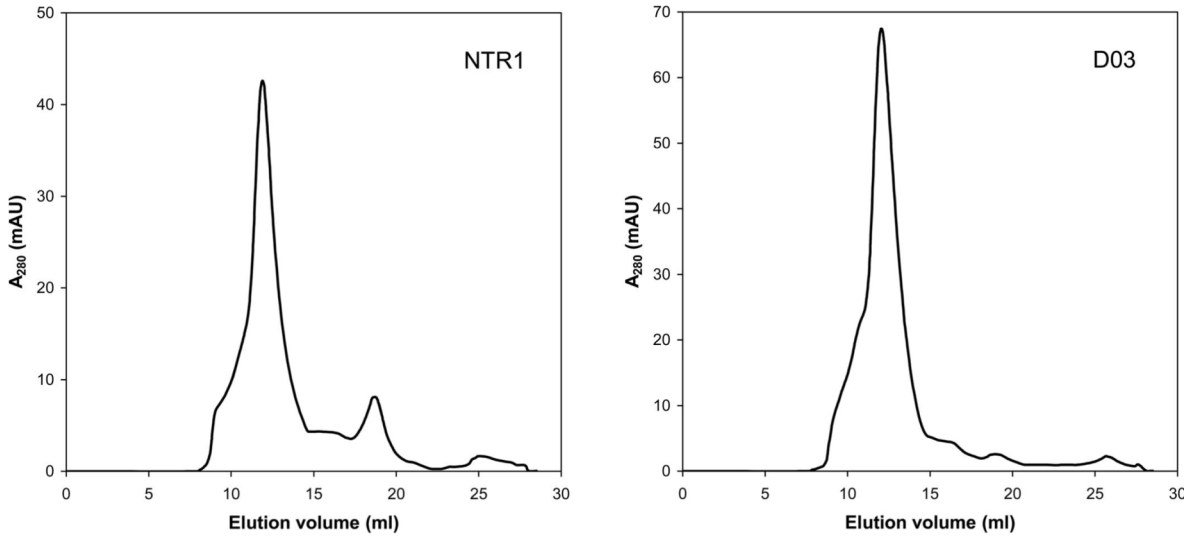


Figure S17: Gel filtration profiles of NTR1 (*Left*; loaded 200 µg of functional receptor at 6.7 µM) and D03 (*Right*; loaded 310 µg of functional receptor at 10.4 µM). The main protein peak elutes at Ve = 11.9 ml and consists of functional receptor (see **Fig. S18**).

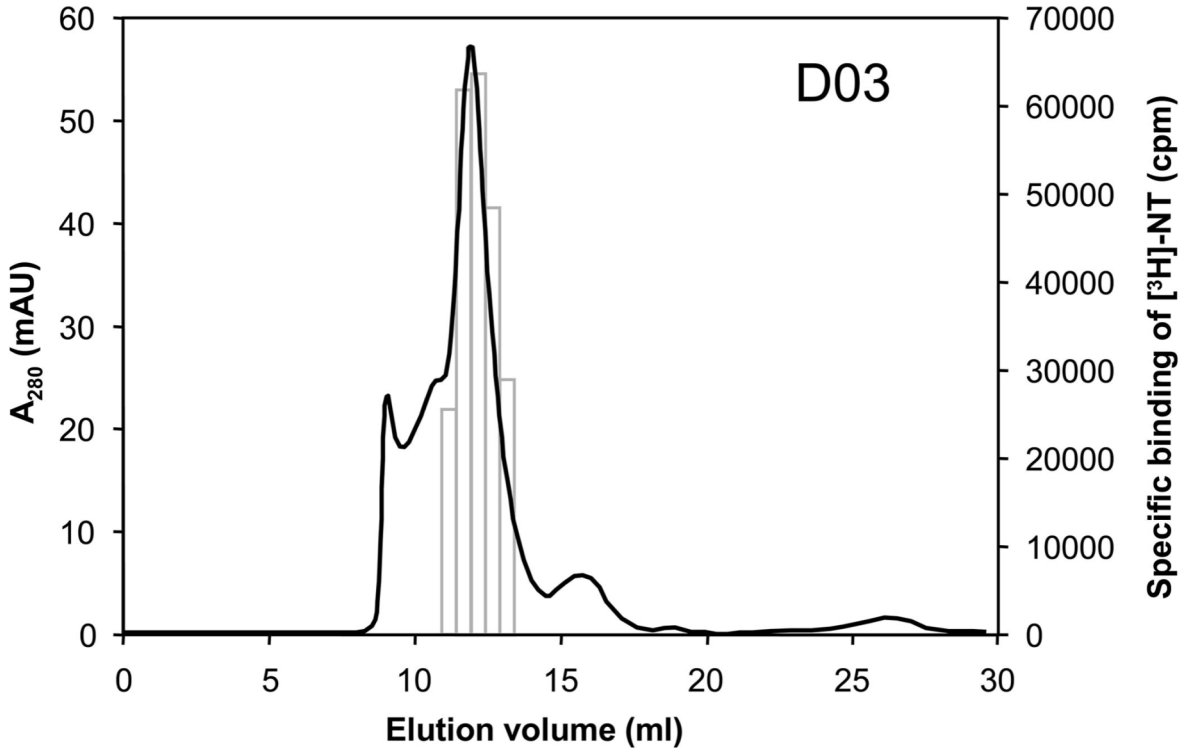


Figure S18: Overlay of a D03 gel filtration profile (solid black line) and the corresponding functional receptor content of collected sample fractions as assessed by [^{3}H]-NT binding assays (open gray columns). The dominant peak (Ve = 11.9 ml) contains functional receptor.

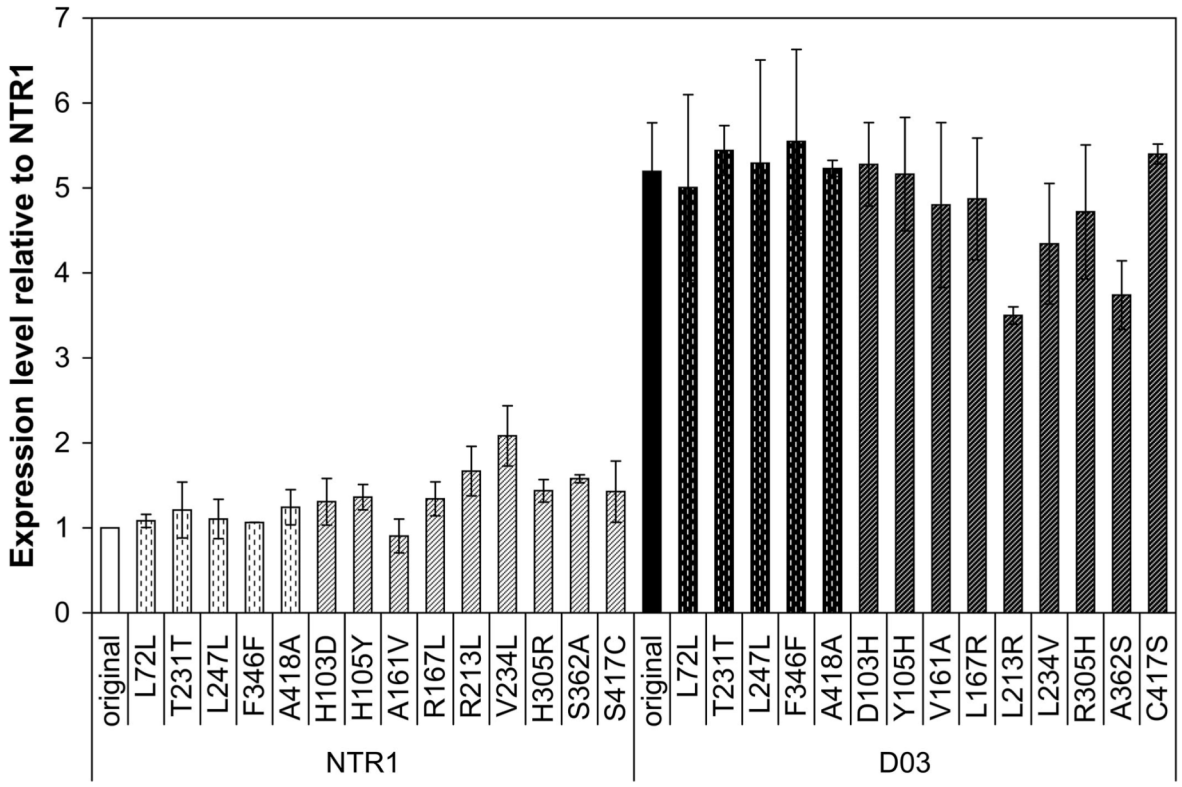


Figure S19: Analysis of individual mutations in D03, which contains 14 nucleotide substitutions compared to NTR1. The first 15 columns in the graph represent NTR1 (solid white), 5 silent substitutions (vertical hash pattern), and 9 single-site mutants of WT that each introduces the evolved amino acid at that position in D03 (diagonal pattern). Likewise, the last 15 columns in the graph represent D03 (black bar), 5 silent substitutions (vertical hash pattern), and 9 single-site mutants of this GPCR that each reintroduces the corresponding WT amino acid (diagonal pattern). All constructs were expressed in the original expression vector lacking the backbone mutation (C7054G). Results represent three independent expression series.

Chapter 5

A generic directed evolution method for producing well-expressed and stable GPCRs in *E. coli*

5.1 Abstract

We recently developed a novel method for the directed evolution of integral membrane proteins in the inner membrane of *Escherichia coli*[1]. Using as the first example the neurotensin receptor NTR1, a mammalian G-protein coupled receptor (GPCR), we arrived at a sequence with a 10-fold increase in functional expression that still retains the biochemical properties of wild type. This mutant also shows enhanced heterologous expression in eukaryotes (12-fold in *Pichia pastoris* and 3-fold in HEK293T cells). Most importantly, the evolved mutant also shows greater stability when detergent solubilized and purified, indicating that the biophysical properties of the protein had been under the pressure of selection. To demonstrate that the approach can be applied generally, we have now studied three additional human GPCRs. Unmodified wild-type receptor cDNA was subjected to this novel membrane protein engineering method. Functional expression could be successively increased with each round of evolution for all three GPCRs. Interestingly, for one receptor – a peptide binding GPCR which to date has resisted expression in *E. coli* – various variants could be quickly evolved to mg/l functional expression levels. In the present study we also developed a new stability screen in a 96-well assay format to quickly screen the pools of evolved receptor variants for candidates showing increased thermal stability in detergent solubilized from. The improvements are based on cumulative small changes in the receptor sequence. They are achieved by a combinatorial approach that does not require the input of rational design principles. Existing roadblocks in structural and biophysical studies of these important targets can now be removed by providing sufficient quantities of correctly folded and stable receptor protein.

5.2 Introduction

G-protein coupled receptors (GPCRs) represent the largest superfamily of cell surface receptors in the living world. They respond to an enormous diversity of extracellular signals, including neurotransmitters, hormones, ions and photons. When an incoming signal molecule or photon interacts with the GPCR, the receptor undergoes large conformational rearrangements whereby the signal is directly transmitted to intracellular G-proteins[2]. When activated by a GPCR, the G-proteins switch from a GDP to a GTP bound state, and the subsequent dissociation into their α- and $\beta\gamma$-subunits can activate a variety of different intracellular signaling pathways.

Because of the enormous diversity in transmitted signals, GPCR signaling is involved in nearly every physiological process, and deregulated signaling readily leads to pathologic conditions. Of the nearly 800 different GPCRs identified in the human genome[3], about 300 GPCRs are considered pharmaceutically relevant (excluding the receptors for odor or taste). Presently, about 30 of these are targeted by about 25% of all known drugs in clinical use. The remaining majority of GPCRs are considered potential high-value drug targets. Due to the central biomedical role of GPCRs in human physiology and pathology GPCRs have attracted a lot of research in academia and industry.

Among these research efforts, the crystallographic determination of a three-dimensional receptor structure represents a central methodology for investigating GPCR function at atomic resolution. Despite tremendous research efforts made in GPCR crystallography, structure determination remains enormously difficult and to date the structures of only five different GPCRs could be solved[4-8]. Moreover, this set is somewhat redundant, because the β_1 and β_2 adrenergic receptors are very closely related and the visual cone pigment rhodopsin was solved from two different species. Yet, the structures provide detailed structural insights into receptor function and ligand binding contacts at atomic resolution and they were successfully used for structure based ligand discovery[9]. Analysis of the five receptor structures reveals many conserved structural features as well as many important structural differences that may account for the enormous functional diversity observed in GPCR signaling. It is very likely that the important and often unique structural differences cannot be extrapolated between different members of the GPCR superfamily. It will therefore be necessary to solve the structure of most GPCRs individually to fully explore their pharmacological potential.

The experimental procedure of obtaining a crystal structure typically involves three steps: the protein has to be functionally expressed, solubilized in detergent micelles, and crystallized. However, for most integral membrane proteins, especially for those of animal origin such as GPCRs, each of these

steps turns out to be very problematic. GPCR expression in native tissue is typically very low and it is therefore necessary to set up a robust heterologous overexpression system. Standard techniques for finding acceptable overexpression conditions include screening of expression hosts, promoters, fusion adducts and protein homologs. However, even if all of these expression parameters are optimized, most GPCRs may not yield sufficient material for crystallographic studies. When acceptable expression conditions are found, the problem of extracting the GPCR out of the lipid bilayer by detergents limits the process towards crystallization, because GPCRs are intrinsically unstable proteins and quickly lose their native fold when solubilized.

In a previous study we addressed these roadblocks by analyzing the capacity of the receptor sequence itself as an experimental parameter for overcoming the problems of low receptor expression and stability[1]. To this end we developed an evolutionary selection method consisting of random mutagenesis and selection by fluorescence-activated cell sorting (FACS) based on the versatile expression host *Escherichia coli*. By applying this evolutionary selection method to the model GPCR neurotensin receptor NTR1 we identified mutations which strongly improve functional surface expression and stability, while wild-type biochemical function was largely retained[1]. In the present study we further validated this selection method on three additional human GPCRs of the rhodopsin family.

More specifically, we wanted to characterize two important technical aspects of the method in order to assess its potential for being generally applicable to the members of the GPCR superfamily. The first aspect relates to the possibility of increasing functional expression independent of the wild-type heterologous expression level, by evolving even very weakly expressed GPCRs. The second aspect relates to the identification of the most thermostable receptor mutants in a pool of well-expressed mutants by a newly developed stability screen in a 96-well assay format. We investigated the performance of the new stability screen to see how readily stable receptors can be identified among the well-expressed ones. By characterizing mutations that increase expression and those that increase stability and those that increase both, important information about the structural basis of those effects can be learned.

5.3 Materials and methods

Library construction. The three human GPCRs to be evolved were the tachykinin receptor TACR1, the α_{1a} adrenergic receptor ADRA1a, and the α_{1b} adrenergic receptor ADRA1b. The wt cDNA encoding these receptors was obtained from the Missouri S&T cDNA Resource Center (www.cdna.org). The full-length cDNA (except the ATG start codon) was cloned into a derivative of the expression vector pRG/III-hs-MBP (kindly provided by R. Grisshammer, National Institutes of Health) via the BamHI and Cfr9I restriction sites. This vector derivative encodes a N-terminal fusion of MBP to the receptor and a C-terminal fusion of TrxA. To create genetic diversity, the DNA encoding the GPCR (excluding the fusion adducts) was amplified by error-prone PCR using the GeneMorph II Random Mutagenesis Kit (Stratagene). Ten ng of template DNA was amplified for 30 cycles of PCR. The randomized PCR product was further amplified by PCR using Phusion polymerase (Finnzyme) to obtain sufficient amounts of DNA for library size ligation into the expression vector pRG. *Escherichia coli* strain DH5α-E (Invitrogen) was used for library cloning and expression. Cells were transformed by electroporation. The final library size was 10^7 to 10^8 transformed clones. Sequencing of single clones of the naïve libraries showed 3-4 nucleotide substitutions per kilobase of DNA for the TACR1 and ADRA1a libraries and 2-3 substitutions for ADRA1b libraries. The sequenced clones showed a homogeneous distribution of mutations, including equal amounts of transitions and transversions, which were equally distributed on the four bases. Clonal amplification of the naïve library after electroporation took place at 37°C in 50 ml 2xYT medium containing 1% glucose. After electroporation the cells were grown for 1 h at 37°C, 100 μg/ml carbenicillin (Carb) was added, and the cells were grown for another 2-3 h. At this point the subpopulation of transformed cells, representing a minor fraction of the total biomass, were mechanically separated from the bulk cell mass of non-transformed cells by passing the liquid culture through a 5 μm filter (Millipore). By this filtration step the transformed cells were recovered in the flow-through, while the non-transformed cells, which fail to divide under Carb selection but keep growing to long aggregating filaments, are retained on the filter. The fraction of transformed cells was centrifuged at 5,000 *g* for 5 min, resuspended in 50 ml fresh 2xYT medium containing 1% glucose and 100 μg/ml Carb, and grown to an OD_{600} of 0.1. The cells were centrifuged at 5,000 *g* for 5 min, resuspended in 1x Hogness modified freezing medium (HMFM) [36 mM K_2HPO_4, 13.2 mM KH_2PO_4, 0.4 mM $MgSO_4$, 1.7 mM Na_3-citrate, 6.8 mM $(NH_4)_2SO_4$, 4.4% (v/v) glycerol], frozen in liquid N_2 and stored at -80°C until further use.

Selection for higher expression. The basic protocol for selecting higher expressing receptor variants has been described in detail[1]. In the present study it was adapted as follows: A frozen library of *E. coli* cells was thawed and grown in 2xYT medium [1% glucose, 100 μg/ml Carb] at 37°C for 2 h. A fresh expression culture in 20 ml 2xYT medium [0.2% glucose, 100 μg/ml Carb] was inoculated at OD_{600} = 0.05, grown at 37°C to an OD_{600} of 0.5 and induced with 0.3 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Expression took place at 20°C for 18 to 24 h. For sorting the libraries by FACS, the cells were diluted into ligand binding buffer at $2 \cdot 10^8$ cells/ml. For cells expressing the ADRA1a or ADRA1b libraries the ligand binding buffer was TKCl [50 mM Tris-HCl, pH 7.4, 100 mM KCl], for TACR1 it was 5xPBS-K [50 mM K_2HPO_4, 9 mM KH_2PO_4, 0.7 M KCl, pH 7.4, (note that all sodium salts were replaced by potassium salts in this buffer)] supplemented with protease inhibitors (Complete Protease Inhibitors EDTA-free, Roche). ADRA1 libraries were labeled with 200 nM prazosin-BODIPY-FL (Invitrogen) for 1 to 2 h on ice. Non-specific binding was determined by the addition of 10 μM unlabeled prazosin (Tocris Bioscience). TACR1 libraries were labeled with 300 nM Substance P Oregon Green (Invitrogen) for 2 to 3 h on ice. Non-specific binding was determined by the addition of 10 μM unlabeled Substance P (Tocris Bioscience). The cells were washed by centrifugation at 10,000 *g* for 90 seconds and resuspension in the same volume of TKCl. Sorting was done on a FACSAria II (BD Biosciences) at 20,000-30,000 events/sec in yield mode (for sorting naïve libraries) or purity mode (for subsequent sorting rounds). The most fluorescent 0.5 to 1.5% of the cells in the population (10^7 to 10^8 cells) were sorted directly into 2xYT (1% glucose, 100 μg/ml Carb) for regrowth and further selection. The most fluorescent cells were enriched by sorting the libraries for 3 to 6 rounds by FACS.

Radioligand binding assay on whole cells. Saturating radioactive ligand binding on whole cells was used to measure the functional surface expression level. Twenty μl of expression culture were added to 160 μl ligand binding buffer in a 96-well polyethylene plate. Ligand binding buffer for ADRA1 clones was 50 mM Tris-HCl, 0.1% (w/v) BSA, 1 mM EDTA, 40 μg/ml bacitracin, pH 7.4 containing 10 nM [^{3}H]-prazosin (Perkin-Elmer). Non-specific binding was determinded by the addition of 10 μM unlabeled prazosin. Ligand binding buffer for TACR1 clones was 5xPBS-K supplemented with 0.1% (w/v) BSA, 40 μg/ml bacitracin and Complete Protease Inhibitors containing 10 nM [^{3}H]-Substance P (Perkin-Elmer). Non-specific binding was determinded by the addition of 10 μM unlabeled Substance P. After incubation for 2.5 to 3 h at 4°C the bound ligand was separated from the unbound by fast vacuum filtration on 96-well filtration plates with glass-fiber filters (MAFBNOB, Millipore) and the filters were

washed with 1 ml ice cold TKCl. The filters were transferred to liquid scintillation vials containing 5 ml Optiphase HiSafe 2 scintillation cocktail (Perkin-Elmer), incubated for 24 h, and scintillation was counted on Betamatic II liquid scintillation counter (Kontron). To derive the amount of receptors/cell we assumed a cell density of 10^9 cells/ml per 1 unit of OD_{600}. For determining equilibrium binding affinities, a dilution series of radioligand was used (0.1–15 nM).

Screening of thermal stability in a 96-well assay format. A new method for screening the stability of solubilized receptors in a 96-well assay format was developed. It consists of four steps. First, the receptor is expressed and biotinylated in vivo. Second, the receptor is solubilized and partially purified by immobilization on streptavidin-coated paramagnetic beads. Third, the receptor is exposed to stability screening conditions that induce receptor unfolding (e.g. heat, detergent, buffer). Fourth, the amount of residual folded receptor is determined by a ligand binding assay (LBA) after exposing the receptor to the screening conditions. The detailed protocol is as follows: Biotinylated receptor was expressed using a derivative of the pRG vector that contains a C-terminal AviTag (GLNDIFEAQKIEWHE, biotinylation on K) instead of a deca-His tag. Under standard expression conditions (as described above) more than 60% of the receptors become biotinylated in vivo without the external supply of components of the *E. coli* biotinylation machinery (i.e. the addition of extra free biotin in the expression medium or co-expression of the *E. coli* biotin ligase BirA. Notably, the omission of extra biotin is very advantageous because free biotin strongly competes with biotinylated receptors for binding to streptavidin during receptor immobilization.). After expession, the biotinylated receptor was solubilized and partially purified by immobilization on straptavidin coated paramagnetic beads as follows: In a 2 ml microfuge tube, a cell pellet corresponding to a 2 ml expression culture was washed once with TKCl and then solubilized in 200 µl solubilization buffer containing 50 mM Tris-HCl, pH 7.4, 200 mM NaCl, 30% (v/v) glycerol, 1 mM EDTA, Complete protease inhibitors, detergents ((dodecyl-β-D-maltopyranoside (DDM), 1% (w/v); 3-[(3-cholamidopropyl)-dimethylammonio]-1-propane sulfonate (CHAPS), 0.5% (w/v); cholesteryl hemisuccinate (CHS), 0.1 % (w/v); from Anatrace). Solubilization took place at 4°C for 2 h with gentle agitation. During the last 30 minutes of solubilization 40 µg/ml deoxyribonuclease I (Roche) and 10 mM $MgCl_2$ were added. Solubilized samples were centrifuged at 20,000 *g* for 20 min to remove cell debris. The supernatant was mixed with 10 µl of Dynabeads MyOne Streptavidin T1 beads (Invitrogen) in a 96-well polyethylene plate and receptor immobilization took place for 2 h at 4°C with gentle agitation. The samples were transferred to a 96-well PCR plate (Bioplastics) and the beads were captured by a magnetic capturing device MPC9000 (Dynal). The beads

were washed twice with 150 µl stability assay buffer [SAB, 50 mM Tris-HCl, 100 mM NaCl, 30% (v/v) glycerol, 1 mM EDTA, Complete Protease Inhibitors, 0.01% (w/v) DDM, 0.5% (w/v) CHAPS, 0.1% (w/v) CHS] and resuspended in a final volume of 120 µl SAB. Thermal stability was measured by exposing one aliquot of immobilized receptor to a specific temperature in a thermocycler (Biometra) for 20 min, a second aliquot was kept on ice. After cooling the heated aliquot on ice all samples were assayed for their content of folded receptor by a ligand binding assay. The immobilized receptor was incubated for 1.5 h on ice with 100 µl ligand binding buffer (LBB) [50 mM Tris-HCl, pH 7.4, 15% (v/v) glycerol, 1 mM EDTA, Complete Protease Inhibitors, 0.01% (w/v) DDM, 0.5% (w/v) CHAPS, 0.1% (w/v) CHS] containing 10 nM radioligand. The bound ligand was separated from the unbound by magnetic capturing of the beads, removing the supernatant, washing the beads once with 150 µl LBB and resuspending the beads in a final volume of 20 µl LBB. The beads were centrifuged at 2,000 *g* for 2 min and resuspended by sonication for 10 sec in a water bath. Scintillation was counted in 200 µl OptiPhase Supermix cocktail (Perkin-Elmer) on a liquid scintillation counter (1450 Microbeta Plus, Perkin-Elmer). The non-specific binding signal was less than 10% of the total signal. The stability index (SI) for a receptor is defined as the ratio between the residual binding signal of the heated receptor sample and the initial binding signal of the sample kept on ice.

Molecular Modeling of the evolved mutations. The crystal structure of the β_2 adrenergic receptor (PDB code 2rh1) was used to analyze the evolved mutations in a structural context. Pymol (DeLano Scientific) was used for molecular modeling and visualization.

5.4 Results

5.4.1 Setup of the selection system for increasing expression and stability

To characterize the performance of our selection method for evolving both increased functional expression and receptor stability of GPCRs, we chose three different human GPCRs of the rhodopsin family. These receptors show very different functional wild-type expression levels: tachykinin receptor TACR1 (~30-80 receptors/cell), α_{1a} adrenergic receptor ADRA1a (~350 receptors/cell) and α_{1b} adrenergic receptor ADRA1b (~2000 receptors/cell). These particular receptors were primarily chosen because they had high affinity fluorescent ligands commercially available. TACR1 is a peptide binding receptor and it binds to the eleven amino acid long peptide Substance P, which acts as an endogenous high-affinity agonist for TACR1. For these selections, Substance P was used as an Oregon Green derivative (SP-OG; **Fig. 5.1**). ADRA1a and ADRA1b both belong to the amine binding family of GPCRs and both receptors bind to the small molecule antagonist prazosin with high affinity. For these selections, prazosin was used as a BODIPY-FL derivative (prazosin-BFL; **Fig. 5.1**).

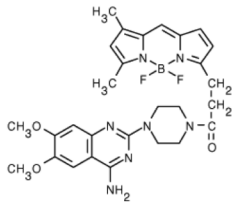
Receptor	Fluorescent ligand	Pharmacophore	Fluorophore	Chemical formula or structure
TACR1	SP-OG	Substance P (SP)	Oregon Green (OG)	RPK^{OG}PQQFFGLM
ADRA1a and ADRA1b	Prazosin-BFL	Prazosin	BODIPY-FL (BFL)	

Figure 5.1: Overview of the fluorescent ligands used for the selections by FACS.

For the selections, we used the highly specific ligand binding fluorescence signal as an indicator for functionally expressed receptor inserted in the inner plasma membrane of *E. coli*[1]. In order to bind to the receptors located in the inner membrane, the fluorescent ligands must be able to partition across the outer membrane (OM) of *E. coli*. A suitable binding buffer must therefore be formulated such that the OM becomes permeable for the ligand molecules

without compromising cell viability. Furthermore, the binding buffer should not contain any components that inhibit ligand binding to the receptor. For TACR1 maximal specific ligand binding of SP-OG was observed in a high-salt buffer containing 5xPBS. The sodium in PBS was substituted by potassium to avoid the inhibitory effect of sodium ions on SP-OG binding. For ADRA1a and ADRA1b maximal specific ligand binding of prazosin-BFL was observed in a buffer containing 50 mM Tris-HCl and 100 mM KCl (TKCl).

5.4.2 Selection of receptor variants showing increased expression

The functional expression level of the evolving receptor libraries increased with each round of mutagenesis and selection as evidenced by an increase in mean specific fluorescence intensity (MFI) (**Fig. 5.2**). The evolution of TACR1 started from a wt receptor that is expressed very weakly as indicated by the histograms of total and non-specific MFI that are almost overlapping for wt TACR1 (**Fig. 5.2**). After four rounds of evolution (ep4 pool, "ep" refers to "error prone PCR", the number "4" refers to four rounds of evolution; one round of evolution consists of one error prone PCR and four to six rounds of sorting by FACS) the specific MFI of the selected TACR1 library pool was approximately five times higher than that of the wt receptor. Similarly, the evolution of ADRA1a started from a relatively weakly expressed wt receptor (**Fig. 5.2**). After four rounds of evolution (ep4 pool) the specific MFI of the selected ADRA1a library was 2.5 times higher than that of the wt receptor. In contrast to the low wt expression levels of TACR1 and ADRA1a, the wt ADRA1b receptor was expressed relatively well, i.e. approximately five times better than its close homolog ADRA1a. After two rounds of evolution of ADRA1b, the selected library (ep2 pool) reached a specific MFI two times higher than that of the wt receptor. ADRA1b was not evolved beyond the ep2 pool.

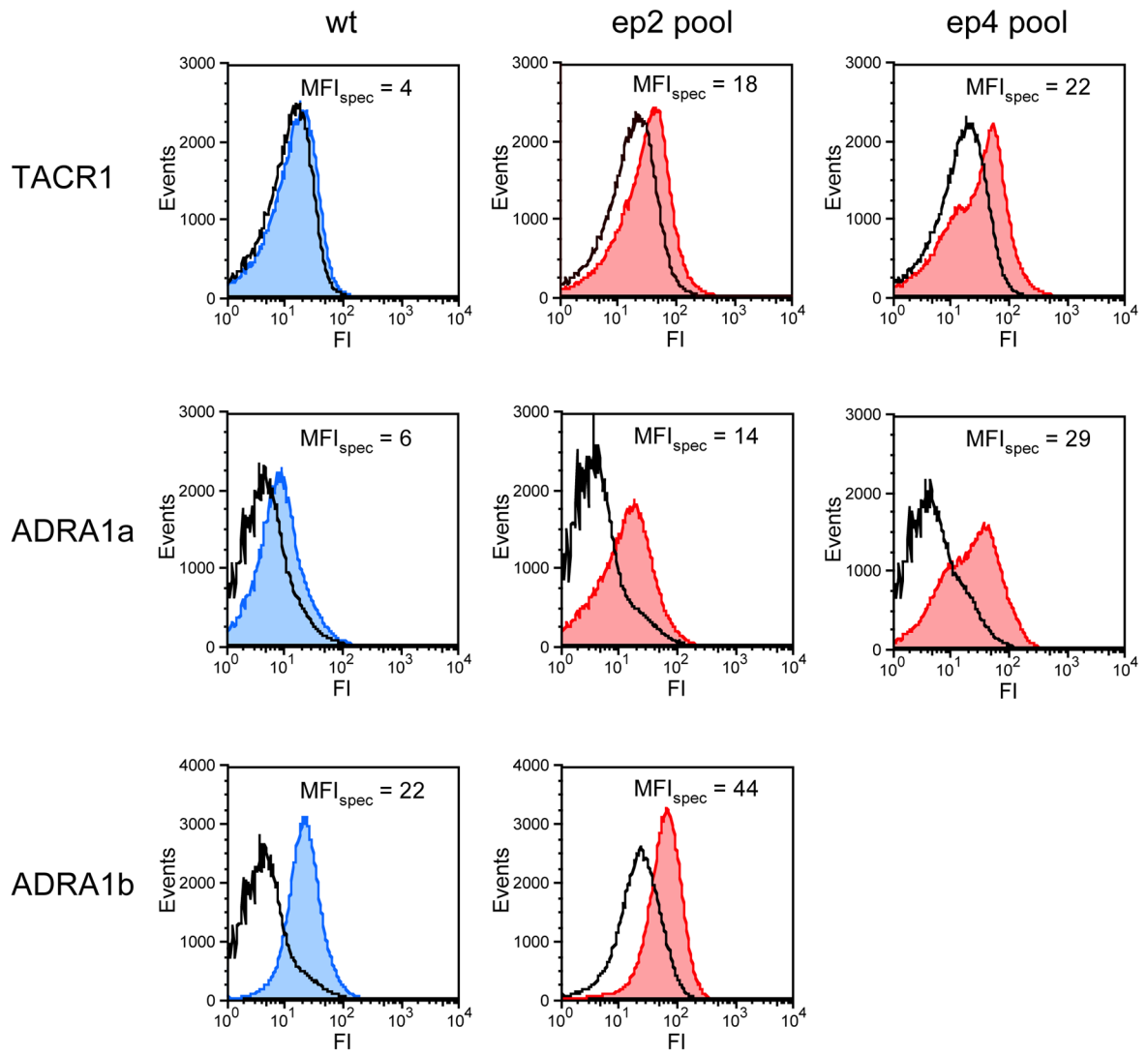


Figure 5.2: Evolution of the expression signal for the three receptors TACR1, ADRA1a, and ADRA1b. The fluorescence distribution is shown for cells expressing the wt receptors (blue histograms) and for pools of evolved cells expressing a collection of mutant receptors (red histograms) after two and four rounds of evolution (ep2 pool and ep4 pool respectively). One round of evolution consists of one error prone PCR followed by four to six rounds of sorting by FACS. ADRA1b was evolved for two rounds only (ep2). The expression signal, given as the specific mean fluorescence intensity (MFI_{spec}), is obtained by subtracting the background MFI (transparent histograms) from the total MFI (colored histograms) of cells labeled with fluorescent ligand. The transparent histograms are obtained by competing the fluorescent ligand with an excess of unlabeled ligand.

To quantitatively assess the increase in expression of single evolved receptor variants within the evolved pools, we performed radioligand binding assays on whole cells of randomly picked single clones.

TACR1: The wt cDNA of TACR1 is very weakly expressed in *E coli*, at a level of 30-80 receptors/cell. In total, we applied four rounds of evolution to TACR1 (four error prone PCRs, each one followed by four to six rounds of sorting by FACS). After two rounds of evolution, we isolated single receptor variants that expressed up to 1200 receptors/cell (**Fig. 5.3a**). Two additional rounds of evolution augmented the expression level further and the best receptor variants expressed approximately 2000 receptors/cell, which represents a 25-fold increase in expression compared to wt TACR1 (**Fig. 5.3a**).

ADRA1a: The evolution of ADRA1a started from a wt cDNA that was expressed at about 350 receptors/cell. We evolved the ADRA1a receptor for 4 rounds (four error prone PCRs, each one followed by four to six rounds of sorting by FACS). After two rounds of evolution we found that the best receptor variants expressed about 10-fold better than wt ADRA1a **(Fig. 5.3b)**. After two additional rounds of evolution, expression increased further and the best clones now expressed more than 6000 receptors/cell, which is about 17-fold better than wt ADRA1a **(Fig. 5.3b)**.

ADRA1b is a close homolog of ADRA1a, which showed a relatively high wild-type expression level of 2000 receptors/cell to begin with. To our knowledge, this represents the highest functional expression level in *E. coli* for any wt GPCR. We included ADRA1b in our set of receptors to be evolved because we wanted to investigate if the expression of a relatively well-expressed GPCR may be further augmented. After two rounds of evolution we isolated clones that expressed 1.5 to 2-fold better than wt ADRA1b (**Fig. 5.3c**). This relatively low increase in expression of ADRA1b compared to the increase observed for the other evolved receptors is very likely related to the problems we encountered in preparing the mutant libraries for ADRA1b. Due to the low amplification efficiency in the error-prone PCR step, the ADRA1b cDNA was randomized very inefficiently. Even after two rounds of randomization, the mutational load of the naïve library was very low (data not shown). To further evolve ADRA1b expression in future studies, it will be necessary to optimize the conditions for error-prone PCR procedure for this particular cDNA.

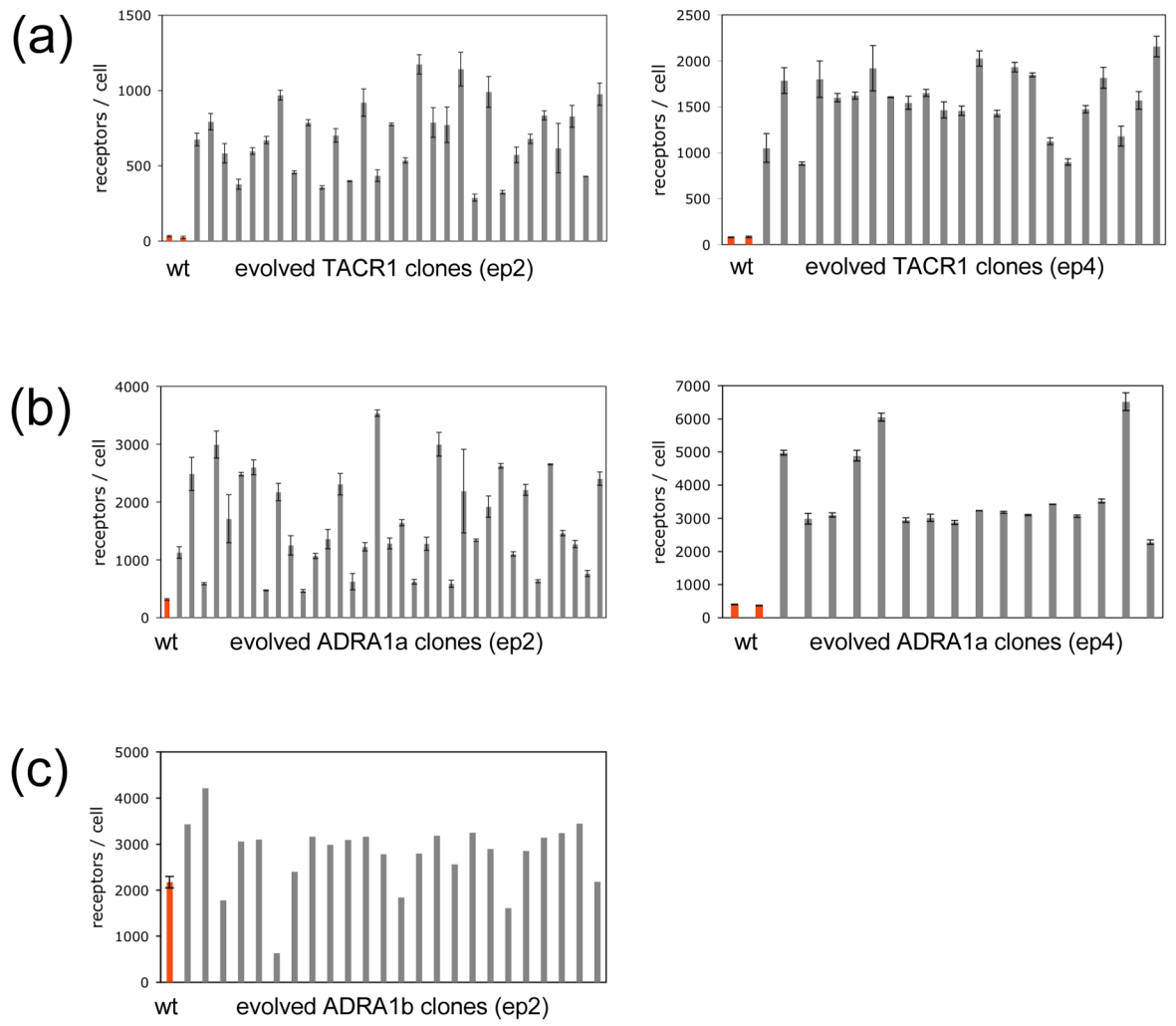


Figure 5.3: Increased functional expression level of evolved receptor variants of TACR1, ADRA1a and ADRA1b after two and four rounds of evolution. One round of evolution consists of one error prone PCR followed by four to six rounds of sorting by FACS. The expression level of single clones was measured by radioligand binding assays on whole cells expressing the evolved receptor variants (grey bars) and the corresponding wt receptors (orange bars). **(a)** Single evolved clones of TACR1 isolated after two rounds (ep2) and four rounds (ep4) of evolution. **(b)** Single evolved clones of ADRA1a isolated after two rounds (ep2) and 4 rounds (ep4) of evolution. (c) Single evolved clones of ADRA1b isolated after two rounds for evolution (ep2). ADRA1b was evolved for two rounds only. The error bars indicate SD of duplicate measurements of a representative experiment.

5.4.3 Identifying receptor variants with increased thermal stability

Establishing high expression levels of a correctly folded and membrane-inserted GPCR is the first critical step in the process of producing sufficient amounts of functional protein for biophysical and structural studies. The second equally important step towards this goal consists of extracting the intact receptor from the lipid bilayer by the help of detergents and maintaining the functional receptor fold in detergent micelles. The parameter that determines the success of this second step most decisively is receptor stability. We were therefore interested to determine how readily we could identify receptor variants with increased stability within the evolved pools of higher expressing clones.

The thermal stability of individual receptor variants was thus determined after detergent solubilization and partial purification. For this purpose, one receptor sample was exposed to a fixed elevated denaturation temperature (defined individually for each of the three receptors) for a fixed period of time (20 minutes), a second sample was kept on ice. Both of the samples were then assayed for their content of folded receptor by a radioligand binding assay (LBA). We define the stability index (SI) for each receptor variant as the ratio between the residual binding signal of the heated sample and the initial binding signal of the sample kept on ice expressed as a percentage. The denaturation temperature was defined as the temperature where the wt receptor would retain approximately 50% of the initial binding signal after heating at this temperature. We found considerable differences in the thermal stabilities of the three wt receptors and we therefore screened the corresponding evolved receptor variants as follows: wt ADRA1a was the most stable wt receptor and it retained approximately 55% binding signal after 20 minutes at 40°C. wt ADRA1b, although 5 times better expressed than its close homolog ADRA1a, was less stable; it retained 20% binding signal after 20 minutes at 36°C. For wt TACR1 we could not determine thermal stability because it was not possible to measure sufficient specific binding signal after receptor solubilization. Instead of wt TACR1, which is too unstable, we therefore decided to use as a reference point the evolved receptor variant T1-C0, which we had isolated in the ep2 pool. We chose this particular clone, because it is well expressed and it deviates from the wt TACR1 sequence by only 4 amino acid substitutions. Clone T1-C0 retained 30% binding signal after heating it for 20 minutes at 31°C.

Because we wanted to perform the stability screen on as many receptor variants as possible, we had to revise the ligand binding assay (LBA), which is the core step of the conventional method for stability assessment. In the conventional method, each sample has to be processed by a small size

exclusion spin column to separate bound from unbound ligand and assess the ligand binding signal, since solubilized receptor cannot be bound to filters. This spin-column step cannot be performed in a 96-well assay format and therefore strongly limits the assay throughput. The key feature of our new method is the immobilization of biotinylated receptor on streptavidin coated paramagnetic beads. By immobilizing the receptor, all essential experimental steps of the stability assay – purification, heat treatment and LBA – can be performed with small receptor amounts and in a highly parallelized format. Immobilized receptor can easily be separated from detergent-solubilized lysates of whole cells by magnetic force, which yields highly concentrated and purified receptor preparations. Most importantly, magnetic capturing also allows for a convenient separation of bound from unbound ligand in the LBA, which avoids the handling of size exclusion spin columns. All essential steps can therefore be performed in a 96-well assay format.

We initially screened 100 to 200 evolved receptor variants for each of the three receptors by applying the new stability screening method. After the initial screen the measurement was repeated with a subset of receptors to confirm the results from the initial screen.

TACR1: We screened receptor variants of the ep4 pool at 31°C and found a broad distribution of stability indices (**Fig. 5.4a**). As mentioned above, it is not possible to compare these data to the stability of wt TACR1, and we compared them instead to an alternative reference point, namely the evolved clone T1-C0. We found that most evolved receptors of the ep4 pool were more stable than T1-C0. Notably, the most stable receptor we isolated (T1-E11) showed a stability index of 91% compared to 31% for T1-C0. To analyze how receptor stability relates to the expression level, we plotted the stability data as a function of receptor expression in **Figure 5.4a**. We found that, although the data are only very loosely correlated, there are many clones that are highly expressed *and* more stable.

ADRA1a: We initially screened 120 single clones of the ep4 pool. However, we did not obtain satisfactory results for this pool, because about half of the sequences showed stop codons in their C-termini, which prevented receptor immobilization on streptavidin-coated magnetic beads via the C-terminal biotin. Interestingly, this subset of receptors showing shortened C-termini (and therefore lacking the C-terminal fusion adduct TrxA-Avi-tag) displayed higher expression levels than the full-length receptors, which explains their enrichment in the ep4 pool. Because of the reduced number of full-length receptors observed in the ep4 pool, we decided to return to the ep2 pool, where the evolved receptors did not show C-terminal shortening. By screening the ADRA1a ep2 pool at 40°C, we found a broad distribution of stability indices, ranging from 34% to 96% (SI of wt ADRA1a was 56%) (**Fig. 5.4b**). The most stable evolved receptor variants showed almost no

reduction in ligand binding after the heat treatment (SI of 96%). Plotting the stability data as a function of expression in **Figure 5.4b** shows that many of the better-expressed receptor variants are also more stable than wt ADRA1a. The correlation between the two variables, however, is very loose. Overall, the results on ADRA1a show that the biophysical properties of a weakly expressed and relatively stable receptor can further be improved by applying the method of random mutagenesis and selection by FACS.

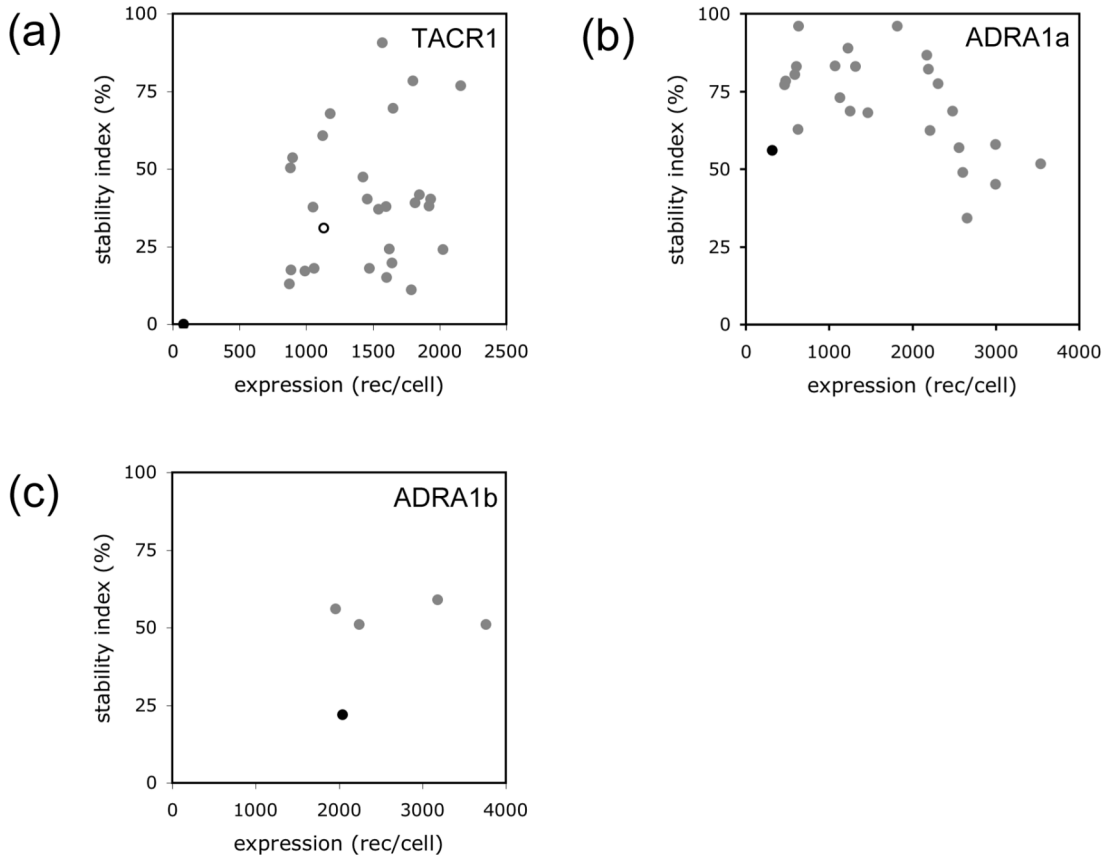


Figure 5.4: Thermal stability as a function of expression level for evolved receptor variants of TACR1, ADRA1a and ADRA1b. To measure thermal stability, the receptors were purified and then heated at a fixed temperature for 20 minutes. The Stability Index (SI) indicates the fraction of receptors that survive the heating step as evidenced by their ability to bind ligand after the treatment. The thermal stabilities of wt receptors (●) are given as a reference point to the evolved receptor variants (●). **(a)** Stability of evolved TACR1 receptor variants of the ep4 pool. The samples were heated at 31°C for 20 min. Wild-type TACR1 failed to be functionally solubilized and its SI is set to 0%. Instead we used the evolved clone T1-C0 (○), which is well expressed and shows only four amino acid substitutions compared to wt, as an alternative reference point for stability. **(b)** Stability of evolved ADRA1a receptor variants of the ep2 pool. The samples were heated at 40°C for 20 min. **(c)** Stability of evolved ADRA1b receptor variants of the ep2 pool. The samples were heated at 36°C for 20 min. The data represent mean values of duplicate measurements from one representative experiment (SD ≤ 10%).

ADRA1b: We screened 120 evolved clones of the ep2 pool. However, only four individual receptor variants could be immobilized on the magnetic beads and assayed for stability. All of them were more stable that wt ADRA1b (SI of 50-60% and 20% respectively) (**Fig. 5.4c**). The majority of receptor variants, however, failed to be immobilized because of the same problem we had observed for clones in the ep2 pool of ADRA1a: the enrichment of stop codons in the evolved pools of both receptors, ADRA1a and ADRA1b, indicates a strong selection in favor of shortened C-termini. To be able to assess the stabilities of the C-terminally shortened receptors, we would have to reclone them into an expression cassette that restores the C-terminal Avi-tag for biotinylation.

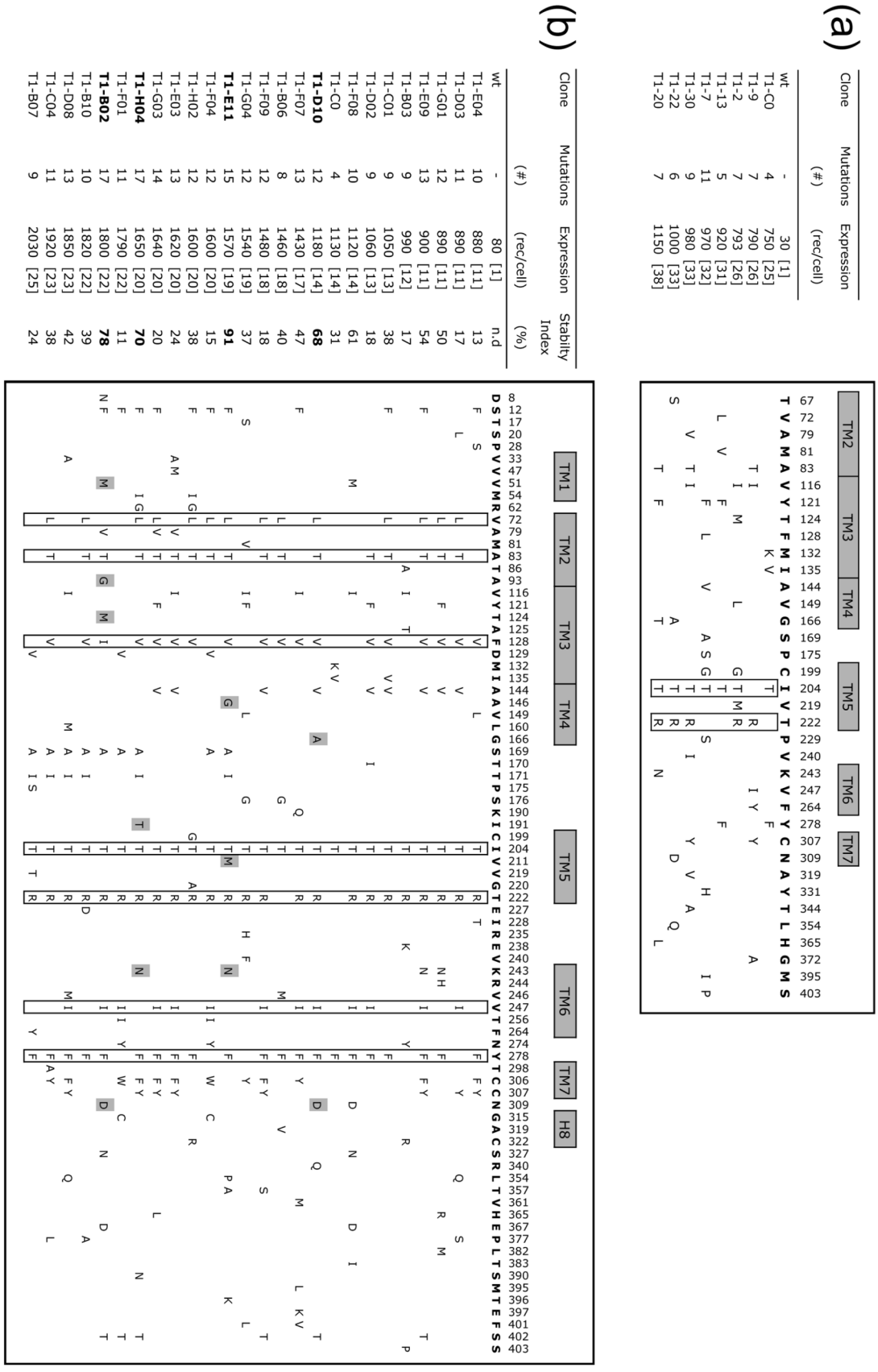
Despite the problems with shortened C-termini observed for ADRA1a and ADRA1b, we have successfully applied the new stability screen to three different wt receptors. For each receptor we could isolate many receptor variants, which showed higher functional expression levels *and* higher thermal stability compared to their wt progenitors.

5.4.4 Relating evolved mutations to increased expression and stability

To find possible structural and mechanistic explanations for how evolved amino acid mutations may influence expression and stability, we mapped them on the recently solved crystal structure of the β_2 adrenergic receptor.

TACR1: To gain insight into mutations that led to the strong increase in expression and stability in the evolution of TACR1, we sequenced single clones of the ep2 and the ep4 pools. After two rounds of evolution (ep2 pool) the most highly expressed sequences show 7 ± 2 (SD) amino acid substitutions per receptor, the sequences are diverse and the mutations are distributed over the entire cDNA (**Fig. 5.5a**). There are two strong consensus mutations, I204T and T222R, both located in TM5 (boxed in **Fig. 5.5a**). Because of their high enrichment in the ep2 pool – the best expressed mutants show at least one of the two mutations – they are likely to be critical for increasing receptor expression. The I204T mutation points into a region surrounded by TM3 and TM4 in the ligand binding crevice. Since it replaces the methyl group of Ile204 by a hydroxyl group, it is tempting to speculate that Thr204 is engaged in an H-bond. The second consensus mutation, T222R, is located at the cytoplasmic end of TM5 and points into the membrane bilayer. The increase in expression may arise because the positive charge of Arg222 may lead to an optimization of the ionic interaction energy between the receptor and the negatively charged phospholipid head groups of the membrane bilayer, and because it satisfies the positive inside rule[10,11]. Besides the two consensus

mutations, there are additional mutations in the ep2 pool, which occur at lower frequencies. These mutations are likely to contribute to increased expression levels in addition to the I204T and T222R consensus mutations.



(a)

Clone	Mutations (#)	Expression (rec/cell)
wt	-	30 [1]
T1-C0	4	750 [25]
T1-9	7	790 [26]
T1-2	7	793 [26]
T1-13	5	920 [31]
T1-7	11	970 [32]
T1-30	9	980 [33]
T1-22	6	1000 [33]
T1-20	7	1150 [38]

(b)

Clone	Mutations (#)	Expression (rec/cell)	Stability Index (%)
wt	-	80 [1]	n.d
T1-E04	10	880 [11]	13
T1-D03	11	890 [11]	17
T1-G01	12	890 [11]	50
T1-E09	13	900 [11]	54
T1-B03	9	990 [12]	17
T1-C01	9	1050 [13]	38
T1-D02	9	1060 [13]	18
T1-F08	10	1120 [14]	61
T1-C0	4	1130 [14]	31
T1-D10	12	1180 [14]	**68**
T1-F07	13	1430 [17]	47
T1-B06	8	1460 [18]	40
T1-F09	12	1480 [18]	18
T1-G04	12	1540 [19]	37
T1-E11	15	1570 [19]	**91**
T1-F04	12	1600 [20]	15
T1-H02	12	1600 [20]	38
T1-E03	13	1620 [20]	24
T1-G03	14	1640 [20]	20
T1-H04	17	1650 [20]	**70**
T1-F01	11	1790 [22]	11
T1-B02	17	1800 [22]	**78**
T1-B10	10	1820 [22]	39
T1-D08	13	1850 [23]	42
T1-C04	11	1920 [23]	38
T1-B07	9	2030 [25]	24

Figure 5.5: Evolved amino acid mutations of receptor variants of TACR1. **(a)** Mutations of the most highly expressed receptor variants of the ep2 pool. **(b)** Mutations of randomly picked receptor variants of the ep4 pool. The receptors are listed based on increasing expression level. The value given in square brackets indicates the factor of expression increase compared to wt expression. For each mutation the corresponding wt amino acid and the position in the TACR1 sequence are indicated on top of the mutation data. Mutations occurring at a frequency of 50% or higher are defined as consensus mutations (boxed). The most thermostable receptors, which are covered in the Results section, are indicated in boldface, and the mutations occurring only in these clones are highlighted as grey boxes.

To analyze the effect of mutations that increase the stability of TACR1, we sequenced clones of the ep4 pool, which had been extensively screened for stability. These clones cover a large range of stability indices ranging from 10% to 91%. There are 11.6 ± 2.8 (SD) amino acid mutations per receptor on average, which were distributed over the entire cDNA (**Fig. 5.5b**). In addition to the two consensus mutations I204T and T222R, which were already present in the ep2 pool, four more mutations are strongly enriched in the ep4 pool (boxed in **Fig. 5.5b**). These mutations seem to be primarily responsible for a general increase in expression rather than in stability because they are not restricted to very stable clones. To gain insight into the strongly stabilizing mutations, we only analyzed mutations (colored in grey in **Fig. 5.5b**) that occur in the 7TM portion (excluding the N- and C-termini) of the most stable TACR1 clones. These clones are:

T1-E11 (SI = 91%)

T1-B02 (SI = 78%)

T1-H04 (SI = 70%)

T1-D10 (SI = 68%)

T1-E11: Clone T1-E11 showed the highest stability index (91%) and it had two mutations, A146G and V211M, which did not occur in any other sequence. Mapping the mutations on the crystal structure of β_2 adrenergic receptor (ADRB2) shows that the A146G mutation is located at the cytoplasmic end of TM3 and that the V211M mutation is located in the middle of TM5. Both mutations are located on the outside faces of the helices and point into the solvent. Clone T1-E11 also features the K243N mutation, which is also present in other stable clones. K243N is located at the cytoplasmic end of TM6 and points towards IL2. In the absence of a detailed atomic model of TACR1, however, it is not evident how the three mutations may contribute to higher stability of TACR1.

T1-B02: Clone T1-B02 was the second most stable clone (SI of 78%) and it figured among the best expressed ones. It shows four mutations, which are likely to increase stability. The A93G and T124M mutations are unique to clone T1-B02 and the V51M and N309D mutations are shared with other stable clones. The T124M mutation introduces a Met residue in TM3 at the interface between TM3 and TM5 and the Met side chain is located near the helix kink introduced by the highly conserved Pro208$^{5.50}$ in TM5. The bulky side chain of Met124 could possibly provide stabilizing hydrophobic contacts in the interface between TM3 and the helix kink in TM5. Interestingly, the T124M mutation in clone T1-B02 and the V211M mutation in the most stable clone T1-E11, although separated in the sequence, are adjacent in the receptor tertiary structure. Both Met side chains point into the same interface region between TM3 and TM5. This suggests that the helical interface around the helix kink introduced by Pro208$^{5.50}$ may play a critical role in the stability of TACR1. Besides the T124M mutation, the other three mutations (V51M, A93G, and N309D) in clone T1-B02 that may be responsible of increased receptor stability are located in different regions of the receptor. It is not obvious how those may increase stability.

The two very stable TACR1 clones mentioned above are succeeded in stability by the two clones T1-H04 and T1-D10, both showing stability indices around 70%. The unique mutations in these clones are I191T in EL2 and G166A at the extracellular end of TM4.

ADRA1a. To analyze the evolved mutations of the ADRA1a selection, we sequenced 48 single clones of the ep2 pool (these clones correspond to the data points presented in **Figure 5.4b**). There were 24 different sequences among the 48 sequenced clones. These sequences are diverse, the mean deviation from the wt sequence is 6.8 ± 2.2 (SD) amino acids per receptor and the mutations are distributed over the entire cDNA (**Fig. 5.6**).

Clone	Mutations (#)	Expression (rec/cell)	Stability Index (%)
wt	-	320 [1]	56
A1A-12	5	460 [1.4]	77
A1A-08	5	470 [1.5]	78
A1A-29	2	590 [1.8]	80
A1A-03	7	600 [1.9]	83
A1A-26	4	620 [1.9]	63
A1A-42	10	630 [2.0]	**96**
A1A-14	5	1070 [3.5]	83
A1A-01	9	1130 [3.5]	73
A1A-22	7	1230 [3.8]	**89**
A1A-11	8	1250 [3.9]	69
A1A-27	7	1310 [4.1]	83
A1A-44	10	1460 [4.6]	68
A1A-05	9	1820 [5.7]	**96**
A1A-09	6	2170 [6.8]	**87**
A1A-31	6	2190 [6.8]	82
A1A-41	6	2210 [6.9]	63
A1A-20	8	2310 [7.2]	78
A1A-06	4	2480 [7.8]	69
A1A-02	7	2560 [8.0]	57
A1A-07	5	2600 [8.1]	49
A1A-43	7	2650 [8.3]	34
A1A-04	7	3000 [9.4]	58
A1A-28	10	3000 [9.4]	45
A1A-23	10	3540 [11.1]	52

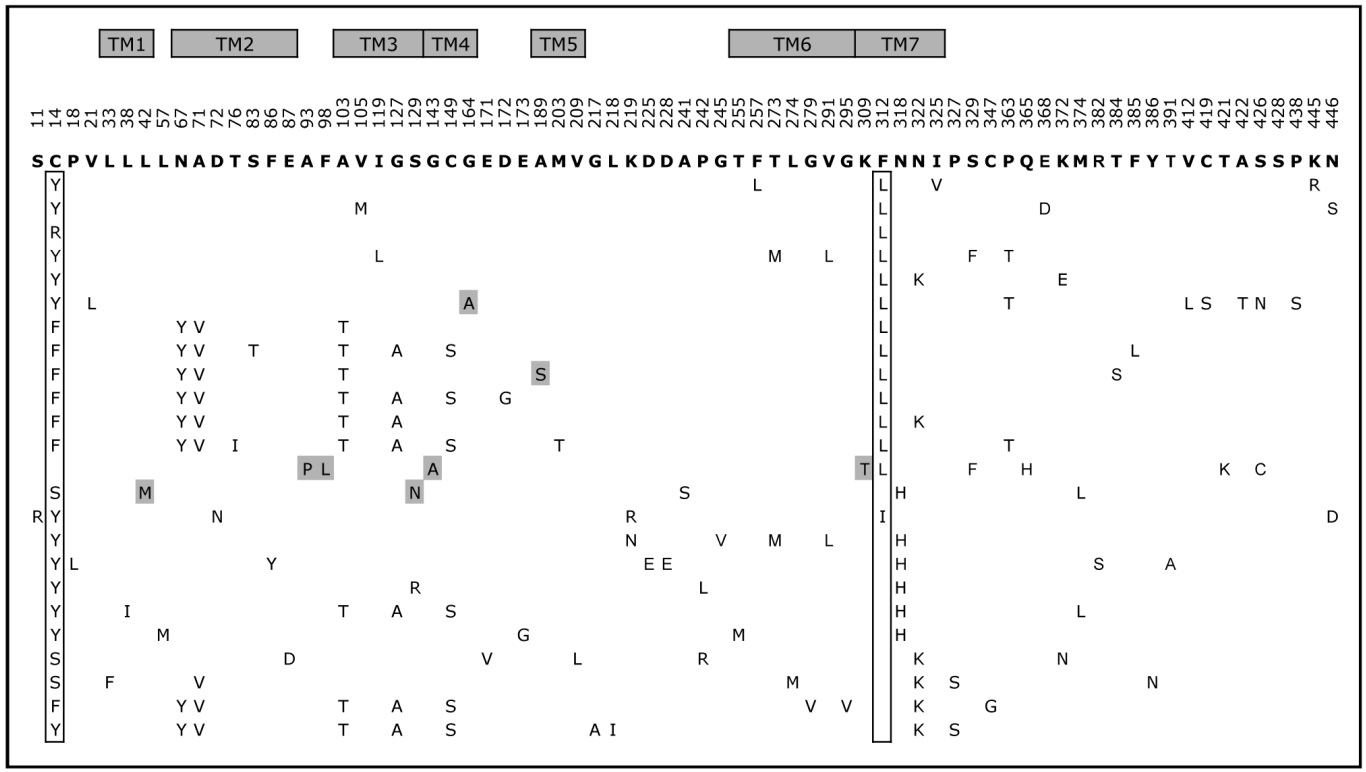


Figure 5.6: Evolved amino acid mutations of receptor variants of ADRA1a. The data show randomly picked clones after two rounds of evolution (ep2 pool). The highlighted elements in the figure are as in **Figure 5.5**.

There are several mutations that occur with high frequency in the evolved ADRA1a sequences. The most prominent of these consensus mutations is found at position 14 in the N-terminus of the receptor. In 96% (23/24) of the evolved sequences the wt Cys residue at position 14 was mutated to either of the amino acids Tyr, Phe, Ser or Arg (**Fig. 5.6**). We analyzed the effect of two different single point mutants at position Cys14 (C14S and C14R) and found that the expression for either of the two mutants was five times higher than for wt ADRA1a (data not shown). This strong effect is most likely related to the elimination of Cys14, which would prevent Cys14 from engaging in the formation of non-specific intramolecular or intermolecular disulfide bonds in the oxidizing periplasm of *E. coli*. The most probable intramolecular partners for such non-specific bonding would be either Cys99 in TM3 or Cys176 in EL2, because they are close to Cys14 in the receptor structure. Interestingly, it is these two Cys residues that form the putative conserved disulfide bond in the native receptor fold. For practically all GPCRs, including ADRA1a, this disulfide is needed for a functional receptor. During receptor expression, the formation of a non-specific disulfide bond involving Cys14 may even be favored over the formation of the specific one, because Cys14 and Cys99 are translocated into the periplasm before Cys176. Interestingly, the closely related receptor subtype ADRA1b shows a Ser residue instead of Cys at the corresponding position. This small difference in the sequence may provide a simple explanation for the 5-fold difference in expression between the two α_1 adrenergic receptor subtypes.

The second strong consensus mutation found in the ep2 pool of ADRA1a is F312L/I located in TM7 and pointing into the ligand binding crevice. The occurrence of F312L/I was associated with two different phenotypic effects. First, all clones that contained this mutation showed a strong increase in stability compared to wt ADRA1a. Second, these clones showed an unexpectedly low binding signal for the antagonist radioligand [^{3}H]-prazosin, which we used for measuring the functional expression level. Both of these effects can be exemplified in the double mutant clone A1a-29, which shows the C14R and F312L mutations. On the one hand, clone A1a-29 was more stable than wt ADRA1a because of the F312L mutation (SI of 83%). On the other hand, A1a-29 showed only a two-fold increase in ligand binding signal instead of a five-fold increase, which we would have expected based on the C14R mutation. While this reduction in ligand binding signal by F312L could be explained by lower receptor expression, it is more likely caused by a reduction in binding affinity for [^{3}H]-prazosin. This interpretation is supported by the observation that Phe312 plays a critical role in the high affinity binding of α_1 adrenergic receptor antagonists[12]. Indeed, we observed that the F312L mutation at the corresponding position in the closely related receptor ADRA1b (the F334L mutation) increased the equilibrium dissociation

constant (K_d) for [^{3}H]-prazosin from 0.3 nM to 2.1 nM (**Fig. 5.7**). A similar drop in affinity in ADRA1a, which is expected as a result of the F312L mutation, would prevent full ligand occupancy of expressed receptors in the radioligand binding assays, which we performed at a concentration of 10 nM [^{3}H]-prazosin. The effective expression level of evolved ADRA1a clones showing the F312L mutation in **Figure 5.6** (and of ADRA1b clones showing the the F334L mutation in **Fig. 5.11**) may therefore be underestimated by roughly 25%. This observation also shows that our selection method does not primarily select for higher ligand affinity, but indeed for the desired properties of expression level and stability.

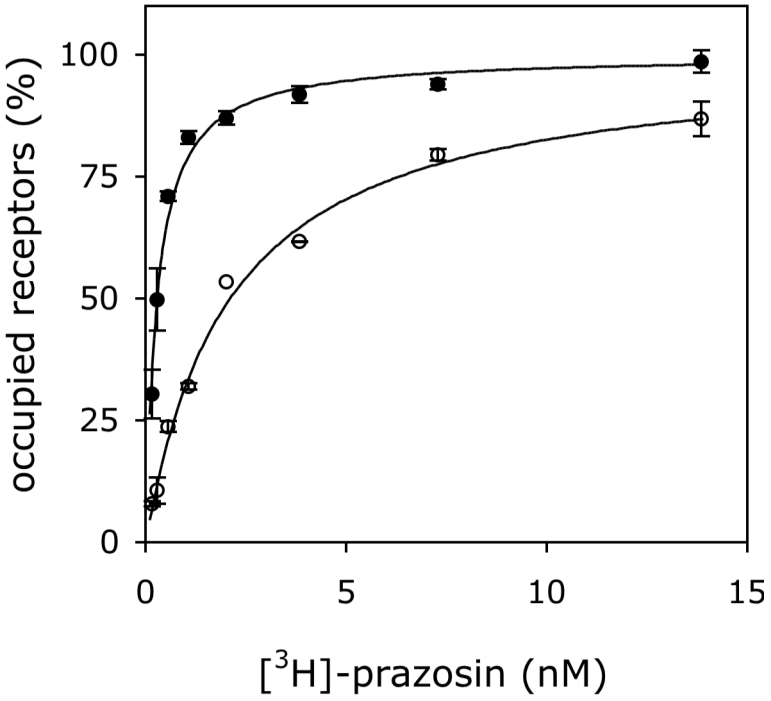


Figure 5.7: Affinity determined with [^{3}H]-prazosin for evolved receptor A1b-C10 compared to wt ADRA1b. The equation $y = y_{max} \cdot x / (K_d + x)$ was used for a non-linear fit of the equilibrium binding data.

Besides the two strong consensus mutations mentioned above (C14X and F312L), two additional consensus mutations (N322K and N318H) occurred in the ep2 pool of ADRA1a, which are associated with the best expressing clones. Molecular modeling using the ADRB2 crystal structure showed that both of these mutant residues are located in TM7 and that they are likely to interfere with the highly interconnected hydrogen bonding network consisting of the conserved residues Asn51$^{1.50}$, Asp79$^{2.50}$, Asn318$^{7.45}$, Asn322$^{7.49}$, Tyr326$^{7.53}$ and several structural water molecules (**Fig. 5.8**). In the model, the introduction of either of the mutations N322K or N318H leads to the elimination of native hydrogen bonds formed by the corresponding wt residues. However, for both mutated residues Lys322 and His318 it is possible

to model rotamers that allow the formation of new alternative polar contacts depicted in **Figure 5.8**. In the case of the N322K mutation, which mutates the Asn residue of the highly conserved NPxxY signature motif in TM7, the mutant Lys322 side chain may possibly form a salt bridge with the highly conserved wt residue Asp79$^{2.50}$ in TM2, which may promote a receptor conformation favoring higher expression

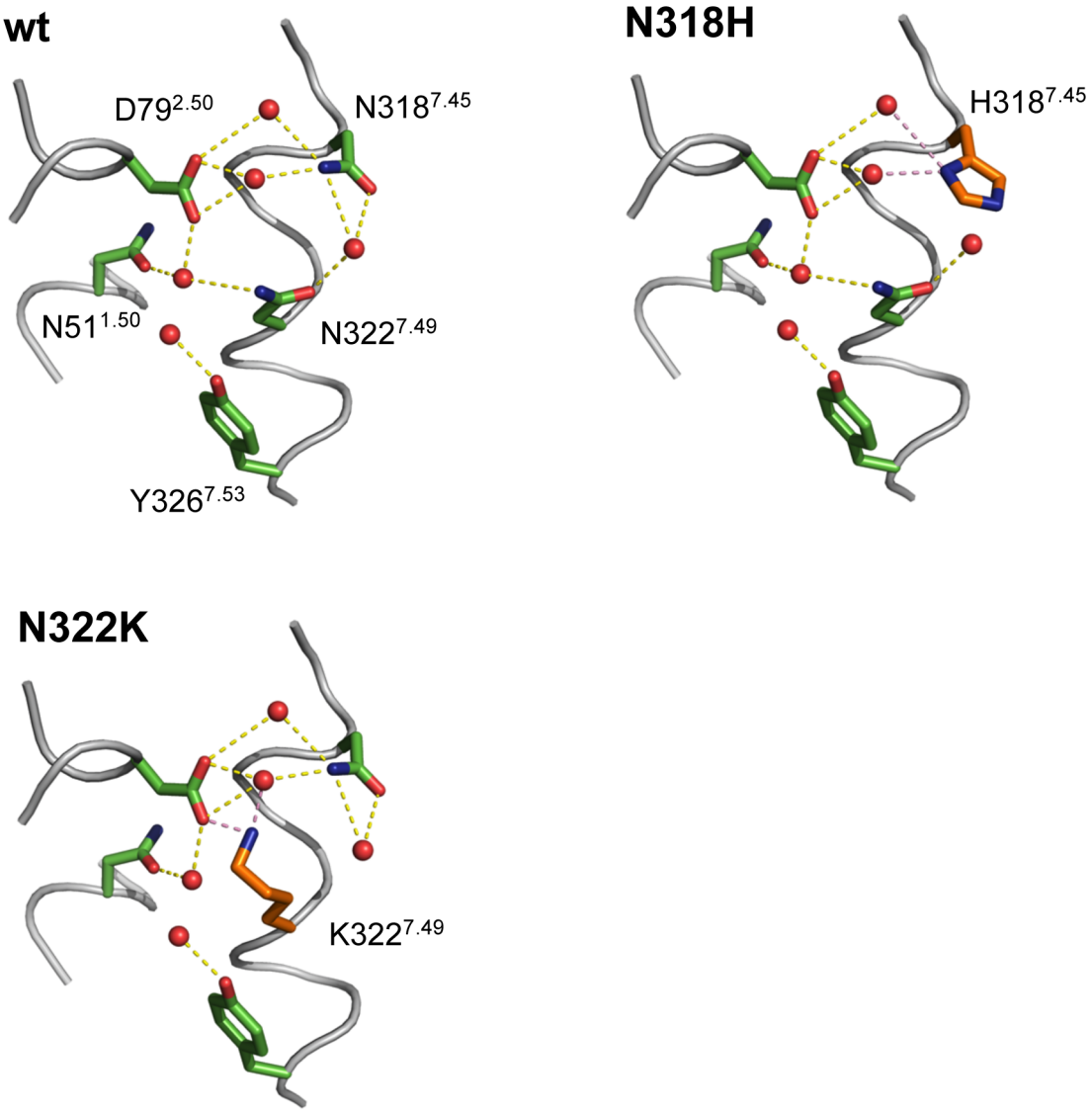


Figure 5.8: Structural model of the N318H and N322K mutations in ADRA1a. These mutations occurred in the highest expressing clones of the ep2 pool of ADRA1a. The model predicts that the mutant amino acids His318 and Lys322 may participate in the highly interconnected hydrogen-bonding network involving highly conserved residues in TM1, TM2 and TM7 and structural water molecules. The putative wt hydrogen bonds are indicated by yellow dots, the hydrogen bonds formed by the mutant residues in pink dots. The model is based on the crystal structure ADRB2 (PDB code 2rh1). Numbers in superscript denote the amino acid position according to the Ballesteros-Weinstein numbering scheme.

In addition to the strong consensus mutations discussed above, we identified mutations in the ep2 pool of ADRA1a, which occurred only the clones that showed the highest thermal stability (**Fig. 5.6**). These clones are:

A1a-05 (SI = 96% ± 1%)

A1a-42 (SI = 96%)

A1a-22 (SI = 89% ± 3.3%)

A1a-09 (SI = 87%).

A1a-05: Of the nine mutations in the most stable clone A1a-05, five are most likely responsible for the high increase in receptor stability because they are located in the 7TM fold. The beneficial effect of one of these mutations (F312L) was already mentioned above. Mapping the remaining four mutations on the crystal structure of ADRB2 provides a possible model for how the interplay of at least three of them (A93P, F98L and K309T) may lead to higher stability of ADRA1a. This model suggests that the Pro residue of the A93P mutation is introduced at a flexible region in the EL1 loop, which may be a critical region for receptor stability (**Fig. 5.9**). Pro93 is flanked by two highly conserved wt aromatic residues (colored in cyan in **Fig. 5.9**), which seem to participate in an extended conserved aromatic cluster around EL1. These flanking residues are Trp92, which forms a tight contact with the conserved disulphide bridge between TM3 and EL2, and Phe94, which contacts conserved aromatic residues in TM2. Replacing Ala for Pro at position 93 would remove flexibility in EL1 and thereby stabilize the highly interconnected aromatic cluster around this loop. In addition to the A93P mutation, the two other mutations F98L and R309T are likely to be coupled to conformational rearrangements induced by Pro93. First, the removal of the bulky benzene ring by the F98L mutation may allow room for the side chain of Phe94 to adopt a new rotamer, which would lead to stabilizing contacts with TM3. Second, the removal of Lys309 by the K309T mutation in TM7, which removes a putative ionic contact with Glu87 in TM2, suggests that TM2 adapts to the changes induced by Pro93, possibly by moving closer to TM7 and forming new contacts. Such a helix movement seems probable because the ionic contact constraining the TM2-TM7 helical interface is lost and because the movement could be facilitated by the helix kink in TM2, which is induced by the conserved Pro81$^{2.50}$. The model further suggests that the conformational rearrangements induced by the mutant residues Pro93, Phe98 and Lys309 could be coupled to the structural changes induced by the stability enhancing consensus mutation F312L in TM7. It is therefore likely that the strong increase in stability in clone A1a-05 arises because of the synergistic action of the all four mutations.

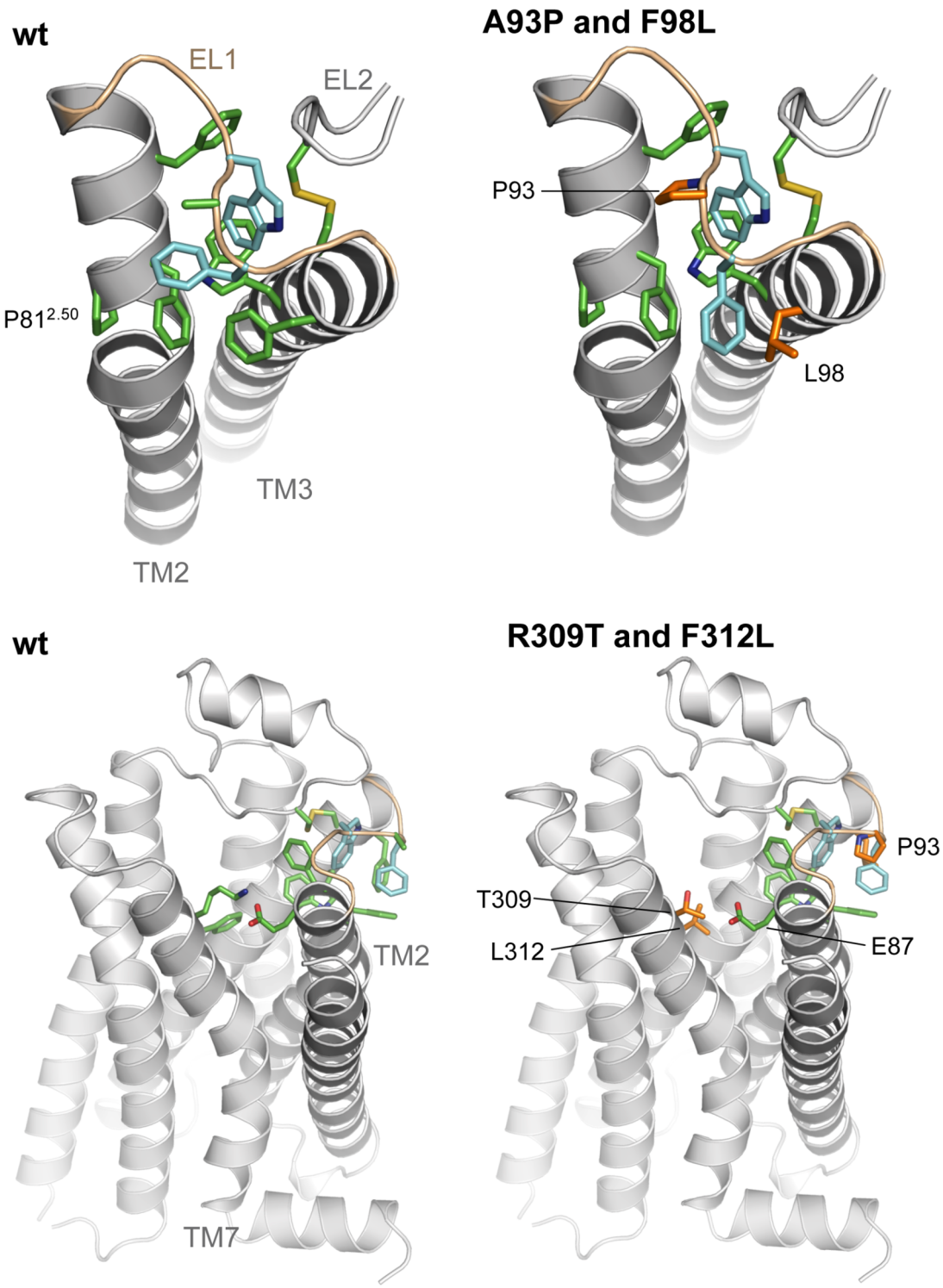


Figure 5.9: Structural model of the mutations A93P, F98L, R309T, and F312L in the highly stabilized clone A1a-05 of ADRA1a. The models on top show an extracellular view of EL1. The models in the bottom show a side view of TM2 and TM7. On the left side are models of the wt receptor and on the right side are models of the mutants (mutant residues are colored in orange). Pro93 is introduced between two highly conserved aromatic residues (colored in cyan), which participate in an extended aromatic cluster around EL1. The model is based on the crystal structure of ADRB2 (PDB code 2rh1).

A1a-42: Clone A1a-42 is another very stable ADRA1a mutant showing a stability index of 96%. In addition to the consensus mutation L312F, it shows two unique mutations in the 7TM fold, V21L and G164A. The V21L mutation lies in the N-terminus and it is difficult to rationalize how it may increase stability. G164A, however, lies at the extracellular helical end of TM4. The substitution of Gly for Ala could possibly remove flexibility and therefore stabilize the connection between TM4 and EL2.

A1a-22: Clone A1a-22 was the third most stable clone (SI of 89%) and its sequence occurred at the highest frequency (7/48 sequenced clones) in the ep2 pool of ADRA1a. It contains the unique A189S mutation in TM5, which is the most likely cause for the increase in stability. Interestingly, all members of the adrenergic receptor family but ADRA1a show Ser and not Ala at the homologous position in TM5. We suppose that the A189S mutation may lead to an increase in stability because the hydrogen-bonding properties of Ser.

A1a-09: Finally, clone A1a-09 appears among the most stable clones (SI of 87%) without featuring the stability enhancing consensus mutation F312L. We speculate that either of the unique mutations L42M or S129N may be responsible for the stabilizing effect. Molecular modeling shows that evolved Met42 in TM1 points into the lipid bilayer. The Met residue may – upon receptor solubilization – possibly engage in favorable hydrophobic interactions with the hydrophobic moieties of detergent molecules. The S129N mutation is located one helical turn away from the Tyr125$^{3.51}$ residue of the highly conserved DRY motif in TM3. The molecular model in **Figure 5.10** shows that the evolved Asn129, but not the wt Ser129, may form a hydrogen bond with Tyr125$^{3.51}$ of the DRY motif in TM3. This interaction may constrain conformational changes induced upon receptor activation, which have been proposed for the receptor region surrounding the DRY motif[13,14].

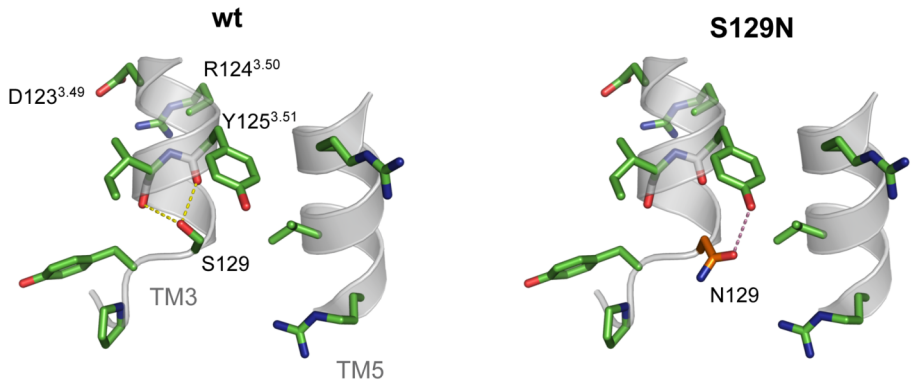


Figure 5.10: Structural model of the S129N mutation in clone A1a-09 of ADRA1a. The wt Ser129 is likely to form hydrogen bonds with the helix backbone and cannot contact Tyr125 of the conserved DRY motif. The mutant Asn129, however, may form a hydrogen bond with Tyr125. This contact may limit conformational flexibility in this important receptor region. The model is based on the crystal structure of ADRB2 (PDB code 2rh1).

ADRA1b. To analyze the mutations leading to increased expression and stability of evolved ADRA1b clones, we sequenced eight selected clones of the ep2 pool (**Fig. 5.11**). Four of them correspond to the only clones that could be immobilized on paramagnetic beads and assayed for stability (**Fig. 5.11**, set A). The other four sequences represent randomly picked clones from the large fraction of clones in the ep2 pool, which were selected for higher expression, but for which we could not to measure stability because they failed to be immobilized on beads (**Fig. 5.11**, set B). On average, the eight clones show 4 ± 2 (SD) amino acid mutations per receptor. Interestingly, all clones show the stabilizing F334L consensus mutation (boxed in **Fig. 5.11**), which is homologous to the F312L mutation observed for many sequences in the ep2 pool of ADRA1a. This Phe to Leu mutation was therefore evolved convergently in the two independent selections of both α_1 adrenergic receptor subtypes. The sequencing data of ADRA1b also suggests that the removal of negative charges in the EL2 loop by mutating either Asp191 or Glu194 to a different amino acid may be responsible for an increase in the expression. Moreover, the removal of a positive charge at the cytoplasmic end of TM6 by the K291N mutation in clone A1b-F08 may be responsible for doubling the amount of expressed receptor.

Clone	Mutations	Expression	Stability Index
	(#)	(rec/cell)	(%)
wt	-	2040 [1]	22
set A			
A1b-B08	4	1960 [1]	56
A1b-C10	3	2240 [1.1]	51
A1b-C02	8	3180 [1.6]	59
A1b-F08	3	3760 [1.8]	51
set B			
A1b-A08	3	n.d	n.d
A1b-A02	3	n.d	n.d
A1b-B02	4	n.d	n.d
A1b-A05	3	n.d	n.d

	TM1					TM6	TM7									
	35	43	187	191	194	291	334	384	386	406	421	467	471	477	481	512
wt	**S**	**T**	**P**	**D**	**E**	**K**	**F**	**C**	**Y**	**S**	**L**	**R**	**G**	**K**	**E**	**N**
set A																
A1b-B08	P	N		H			L									
A1b-C10			L				L									H
A1b-C02				Y	G		L	S	F	*	M		C	R		
A1b-F08						N	L					H				
set B																
A1b-A08					V		L	*							D	
A1b-A02				Y			L	*							D	
A1b-B02				Y	G		L	*							D	
A1b-A05				Y	G		L	*								

Figure 5.11: Evolved amino acid mutations of receptor variants of ADRA1b after two rounds of evolution (ep2 pool). The sequences are split in two parts. (set A) Full-length receptors. (set B) Receptors showing an opal stop codon in their C-terminus. Note that the F334L consensus mutation in ADRA1b corresponds to the F312L consensus mutation in ADRA1a. The highlighted elements are as in **Figure 5.5**.

The observation that many of the evolved ADRA1b clones failed to be immobilized is explained by the fact that the codon for Cys384 near the C-terminus was mutated to an opal stop codon in these receptors. The opal stop codon leads to premature termination of translation and therefore prevents biotinylation of the C-terminal Avi-tag, which is necessary for immobilizing the receptor on streptavivdin-coated beads. Note that clone A1b-C02

(**Fig. 5.11**) could be immobilized despite showing a stop codon. In this case, however, it is not an opal but an amber stop codon, which is strongly suppressed in the *E. coli* DH5-α strain we were using. The amber suppression allows for the translation of a full-length receptor and therefore clone A1b-C02 could be immobilized and assayed for stability. A more general implication of the high occurrence of stop codons in the evolved clones of both α_1 adrenergic receptor subtypes is that shortening their relatively long C-termini is beneficial for high level expression. Furthermore, the observation that receptors became shortened near their C-termini, but never within their 7TM portion, may be explained by the fact that our selection method is based on ligand binding as a stringent selection criterion for conserving the structural integrity of the 7TM fold.

5.5 Discussion

The inability to express GPCRs at high levels and to maintain their stability in detergent micelles restricts the process of crystallization and structure determination. To overcome these restrictions we have developed an evolutionary selection method that allows the rapid identification of mutations in the receptor sequence, which lead to higher functional expression and higher stability. In a previous study we have demonstrated the feasibility of the method on the model protein neurotensin receptor NTR1[1]. In the present study we further validate the method by successfully applying it to three additional human GPCRs of the rhodopsin family. More specifically, we wanted to test if GPCRs could be evolved for higher expression independent of their wild-type expression level, and how readily we could isolated well-expressed mutants, which also show increased biophysical stability in detergent solubilized form.

5.5.1 Increasing expression and stability

This study shows that the functional expression level of three wt GPCRs could be strongly increased independent of their basal heterologous wt expression level. The increase in expression was especially high for the evolved receptor variants of TACR1 and ADRA1a, which both show very low wt expression. For TACR1, which is hardly expressed in *E. coli*, we could improve the expression level 25-fold (from 80 to 2000 rec/cell). For the low expressing ADRA1a we found an 11-fold increase in expression (from 330 to 3500 rec/cell). Even for the relatively well expressed wt ADRA1b, we isolated receptor variants showing a 2-fold increase in expression (from 2000 to 4000 rec/cell). These levels would roughly correspond to 2 mg/l functional GPCR (excluding the fusion partners) at a cell density of OD_{600} = 5, and therefore constitute amounts clearly in the range of structural studies.

The rapid identification of the most stable receptor variants among the well-expressed ones was enabled by a newly developed stability screen in a 96-well assay format. By this screen we isolated highly stabilized and well-expressed clones such as T1-E11 and T1-B02 derived from TACR1 or the clones A1a-05 and A1a-09 derived from ADRA1a. In general, we found that most receptors that were better expressed were also more stable, but the two variables were only loosely correlated. This observation is consistent with the fact that during the selection, mutants are primarily selected for their increased functional expression level rather than biophysical stability. However, the data shows that increased biophysical stability is an important factor for higher expression among several other factors, which are unrelated to receptor stability. Importantly, for many receptors, both properties need to be improved.

A receptor, which cannot be expressed, is not useful for structural studies, even if it is more stable than the wt receptor. Our method allows to quickly find stabilizing mutations, among a pool of better expressed mutants.

5.5.2 Technical aspects of the selection method

In addition to our two primary findings discussed above, a note on three further technical aspects may be of interest when discussing the performance of the presented method. First, the method was implemented without making a priori assumptions about the structure or the function of the targeted receptor. To initiate the selection process we simply chose wt cDNA of receptors, for which fluorescent ligands were commercially available. Genetic manipulations were restricted to the removal of internal restriction sites and did not change the amino acid sequence. The wt receptors were cloned into the expression vector including all of their potentially problematic features such as loops, N- and C-termini or possible protease sites. To initiate the selection process, it was not necessary to employ classical techniques of expression optimization. The results in this study were obtained by keeping all expression conditions constant.

The second technical aspect addresses the concept of introducing mutations into the receptor for increasing expression and stability. In order to avoid the enrichment of highly expressed but misfolded mutants we are employing a selection mechanism that relies on *ligand binding* as a functional selection criterion. The ability to bind a ligand molecule represents a strong criterion for structural integrity because for most GPCRs the ligand binding site is contributed by residues from different parts of the receptor, which must be in a particular three-dimensional arrangement. Therefore, the helices and loops must be oriented in a native-like conformation to provide the functional contacts for ligand binding. Indeed, the evolved receptors typically retain a functional ligand binding site. Radioligand binding analysis showed that the evolved receptor variants bound their cognate ligand with high affinity. As an exception to this general outcome we observed that many evolved clones of ADRA1a and ADRA1b showed the Phe to Leu consensus mutation in TM7. This mutation led to a 10-fold decrease in affinity for prazosin binding to ADRA1b. The loss of affinity, however, coincided with a considerable gain in receptor stability.

We are well aware of the fact, that selecting for mutations that preserve the ligand binding site does not rule out the possibility of selecting for mutations that affect other important functional features of a GPCR, such as functional coupling to G-proteins. In fact it is very probable that some stabilizing mutations do exert their effects by constraining conformational flexibility whereby the equilibrium is shifted towards a more active or more inactive

receptor state. In a previous study on the directed evolution of NTR1, we have analyzed in great detail the signaling profile of a receptor variant (D03), which we had evolved for higher expression and stability[1]. Indeed, the D03 receptor retained wt-like affinity for its agonist neurotensin (as we had expected based on the selection pressure). Its efficacy in Ca^{2+} signaling was still present, but it required a 10-100 fold higher agonist concentration than in wt NTR1, suggesting a higher rigidity of the evolved receptor.

Despite the fact that stabilizing mutations may change the signaling profile of a mutated GPCR we believe that restraining conformational flexibility represents a decisive advantage for the crystallographic determination of GPCR structures. This was recently demonstrated by solving the crystal structure of the turkey β_1 adrenergic receptor, which was made possible by introducing 6 stability enhancing mutations in the receptor[6]. Similarly, the thermal stabilization of bovine rhodopsin by an engineered disulfide bond in bovine rhodopsin has enabled its recombinant production for the first time, as reported in a recent study[15]. This protein could then be crystallized and its structure solved. Before this mutation was found, rhodopsin had to be extracted from the bovine retina to obtain sufficient protein for structural studies[4].

As a third technical aspect of the method, we want to note that the successful selection for higher expression and stability relies on both a strong functional selection criterion (such as ligand binding) as well as on the ability to sample a large sequence space. Sampling the largest sequence space possible increases the chances for isolating the few beneficial mutants among the majority of non-functional ones, which inevitably dominate a naïve gene library generated by random mutagenesis. In this study we routinely analyzed 10^7 to 10^8 mutant clones per selection round. These numbers were achieved on a high speed FACS, which analyzes cellular parameters (such as receptor expression) individually for each clone. Our results indicate that the sequence space was large enough to sample several different amino acids at every given position in the sequence. For example, we found that Cys14 in ADRA1a, a residue that strongly impairs expression in wild-type, was mutated to four different amino acids in the evolved clones (Tyr, Phe, Ser and Arg). Therefore, the Cys codon in the naïve gene library must have been randomized at all three positions of the codon to give rise to the selected mutations. A further evidence for thorough sampling of the sequence space is the convergent evolution of the Phe to Leu consensus mutation in TM7 in two independent selections on the two homologous receptors ADRA1a and ADRA1b. Most importantly, the sampled sequence space was large enough to allow for the identification of very rare but highly beneficial mutations.

5.5.3 Relating mutations to evolved properties

What mutations make the evolved receptors better expressed and more stable and what molecular mechanisms may be involved in the improvements? The general conclusion we draw from this study is that there exist many different mutation patterns for a GPCR sequence that lead to improved expression and stability. For example, we identified mutations that increase expression, but not stability, and most likely do not alter the native tertiary receptor structure at all. This is the case for the mutation of Cys14 in the N-terminus of ADRA1a. Replacing Cys14 increases expression about five-fold without affecting stability. The simplest and most likely explanation for this strong effect may be that the elimination of Cys14 prevents the formation of non-native intramolecular disulfide bonds during co-translational receptor insertion into the membrane. Other mutations, such as the combination of four mutations in the evolved clone A1a-05 of ADRA1a, strongly increase receptor stability, which is the most likely explanation for the increase in expression, too. A detailed examination of a subset of evolved mutations we predict as the most relevant ones is given in the Results section. As yet, many of these predictions remain speculative to a certain degree and it will be necessary to test the mutations individually for their actual contributions. This information would then enable the engineering of receptor variants that would profit from combining the best mutations in a single sequence.

5.5.4 Conclusions

In recent years, the field of membrane protein engineering has advanced more and more. Some of the recently developed techniques aim at improving membrane protein expression level, for example by manually screening blots of random mutants[16]. Other techniques aim at identifying mutations that increase protein stability in detergent micelles. These techniques include the functional screening of random mutants[17] or the systematic screening of alanine substitutions[18]. Yet other techniques aim at identifying functionally critical amino acids in GPCRs[19-21]. Complementary to such approaches, we have developed a protein engineering platform that allows the simultaneous identification of well-expressed and stable GPCR variants that retain their functional ligand binding properties. The method has now been validated on four different GPCR of the rhodopsin familiy. The results in this study show that our method represents a nearly assumption-free experimental platform and it should therefore be generally applicability to other members of the GPCR superfamily. Moreover, the analysis for the selected mutations will be valuable in understanding the structural contribution to stability and folding in GPCRs.

5.6 References:

1. Sarkar, C.A., Dodevski, I., Kenig, M., Dudli, S., Mohr, A., Hermans, E. & Plückthun, A. **Directed evolution of a G protein-coupled receptor for expression, stability, and binding selectivity**. *Proc Natl Acad Sci U S A* **105**, 14808-14813 (2008).
2. Hepler, J.R. & Gilman, A.G. **G proteins**. *Trends Biochem Sci* **17**, 383-387 (1992).
3. Schioth, H.B. & Fredriksson, R. **The GRAFS classification system of G-protein coupled receptors in comparative perspective**. *Gen Comp Endocrinol* **142**, 94-101 (2005).
4. Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M. & Miyano, M. **Crystal structure of rhodopsin: A G protein-coupled receptor**. *Science* **289**, 739-745 (2000).
5. Rasmussen, S.G., Choi, H.J., Rosenbaum, D.M., Kobilka, T.S., Thian, F.S., Edwards, P.C., Burghammer, M., Ratnala, V.R., Sanishvili, R., Fischetti, R.F., Schertler, G.F., Weis, W.I. & Kobilka, B.K. **Crystal structure of the human beta2 adrenergic G-protein-coupled receptor**. *Nature* **450**, 383-387 (2007).
6. Warne, T., Serrano-Vega, M.J., Baker, J.G., Moukhametzianov, R., Edwards, P.C., Henderson, R., Leslie, A.G., Tate, C.G. & Schertler, G.F. **Structure of a beta1-adrenergic G-protein-coupled receptor**. *Nature* **454**, 486-491 (2008).
7. Jaakola, V.P., Griffith, M.T., Hanson, M.A., Cherezov, V., Chien, E.Y., Lane, J.R., Ijzerman, A.P. & Stevens, R.C. **The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist**. *Science* **322**, 1211-1217 (2008).
8. Murakami, M. & Kouyama, T. **Crystal structure of squid rhodopsin**. *Nature* **453**, 363-367 (2008).
9. Kolb, P., Rosenbaum, D.M., Irwin, J.J., Fung, J.J., Kobilka, B.K. & Shoichet, B.K. **Structure-based discovery of beta2-adrenergic receptor ligands**. *Proc Natl Acad Sci U S A* **106**, 6843-6848 (2009).
10. von Heijne, G. & Gavel, Y. **Topogenic signals in integral membrane proteins**. *Eur J Biochem* **174**, 671-678 (1988).
11. White, S.H. & von Heijne, G. **The machinery of membrane protein assembly**. *Curr Opin Struct Biol* **14**, 397-404 (2004).
12. Waugh, D.J., Gaivin, R.J., Zuscik, M.J., Gonzalez-Cabrera, P., Ross, S.A., Yun, J. & Perez, D.M. **Phe-308 and Phe-312 in transmembrane domain 7 are major sites of alpha 1-adrenergic receptor antagonist binding. Imidazoline agonists bind like antagonists**. *J Biol Chem* **276**, 25366-25371 (2001).
13. Rovati, G.E., Capra, V. & Neubig, R.R. **The highly conserved DRY motif of class A G protein-coupled receptors: beyond the ground state**. *Mol Pharmacol* **71**, 959-964 (2007).
14. Greasley, P.J., Fanelli, F., Rossier, O., Abuin, L. & Cotecchia, S. **Mutagenesis and modelling of the alpha(1b)-adrenergic receptor highlight the role of the helix 3/helix 6 interface in receptor activation**. *Mol Pharmacol* **61**, 1025-1032 (2002).
15. Standfuss, J., Xie, G., Edwards, P.C., Burghammer, M., Oprian, D.D. & Schertler, G.F. **Crystal structure of a thermally stable rhodopsin mutant**. *J Mol Biol* **372**, 1179-1188 (2007).
16. Molina, D.M., Cornvik, T., Eshaghi, S., Haeggstrom, J.Z., Nordlund, P. & Sabet, M.I. **Engineering membrane protein overproduction in Escherichia coli**. *Protein Sci* (2008).

17. Zhou, Y. & Bowie, J.U. **Building a thermostable membrane protein**. *J Biol Chem* **275**, 6975-6979 (2000).
18. Tate, C.G. & Schertler, G.F. **Engineering G protein-coupled receptors to facilitate their structure determination**. *Curr Opin Struct Biol* **19**, 386-395 (2009).
19. Li, B., Scarselli, M., Knudsen, C.D., Kim, S.K., Jacobson, K.A., McMillin, S.M. & Wess, J. **Rapid identification of functionally critical amino acids in a G protein-coupled receptor**. *Nat Methods* **4**, 169-174 (2007).
20. Beukers, M.W. & Ijzerman, A.P. **Techniques: how to boost GPCR mutagenesis studies using yeast**. *Trends Pharmacol Sci* **26**, 533-539 (2005).
21. Sommers, C.M. & Dumont, M.E. **Genetic interactions among the transmembrane segments of the G protein coupled receptor encoded by the yeast STE2 gene**. *J Mol Biol* **266**, 559-575 (1997).

General conclusions and future perspectives

The promise of future research in structural biology of GPCRs includes elucidating the differences between the active and inactive receptor conformations, the molecular interactions between receptors and ligands, and structures of GPCRs in complex with effector proteins such as G-proteins or β-arrestins. Therefore, it is necessary to overcome the experimental roadblocks in producing sufficient amounts of soluble, stable and pure GPCRs. In this doctoral thesis I have worked on three different protein engineering strategies for removing some of the largest experimental roadblocks.

The possibility of engineering a water-soluble analog of a GPCR is still a matter of debate and it surely demands a large portion of imagination to believe in its feasibility. A proof of this concept has not yet been demonstrated within the scope of this thesis. Nevertheless, the combinatorial methods of directed evolution may provide the key technology to reach this ambitious goal. To this end I have synthesized a semi-rational gene library that can be used to select for water-soluble GPCR analogs. The next experimental steps will focus on the implementation of a suitable display method (e.g. ribosome display) for performing the selections on the library.

From our study on the engineering of opioid receptors by domain swapping I conclude that small biophysical improvements can be attained by this strategy. However, the improvements are in no way sufficient to produce crystallization-quality soluble material. Nevertheless, the application of this strategy to other GPCR families may be helpful to elucidate structure-function relationships between closely related homologs.

The studies on the directed evolution of GPCRs by FACS were the most successful. We could show that even small changes to the amino acid sequence can lead to strong improvements in functional expression and stability of GPCRs while wild-type biochemical properties can be largely retained. For the rapid identification of critical amino acids we have developed a powerful method based on the principles of directed evolution. Notably, the method represents a nearly assumption-free way of engineering the biophysical properties of integral membrane proteins. Our results on four different GPCRs demonstrate that the method should be generally applicable to the GPCR superfamily. The method, however, is not limited to GPCRs. Any integral membrane protein, for which a specific fluorescent ligand can be synthesized, can potentially be evolved. If the evolved proteins are more stable because of enhanced rigidity, they may also be more likely of form diffraction-quality crystals for structural

determinations. The ability to obtain such high-resolution structures may help to elucidate the molecular basis for activation, inactivation, or pathology associated with a given receptor, and may also provide templates for drug design.

Appendix

Abbreviations

2xYT	double strength yeast extract trypton medium
A_{2A}	adenosine receptor A_{2A}
ADRA1a	α_{1a} adrenergic receptor
ADRA1b	α_{1a} adrenergic receptor
ADRB2	β_2 adrenergic receptor
AU	absorbance unit
BFL	BODIPY fluorescein like
BODIPY	Boron-dipyrromethene
BSA	bovine serum albumin
Carb	carbenicillin
CB_1	cannabinoid receptor type 1
CB_2	cannabinoid receptor type 2
cDNA	complementary DNA
CHAPS	3-[(3-cholamidopropyl)-dimethylammonio]-1-propane sulfonate
CHS	cholesteryl hemisuccinate
D_2	dopamine receptor D_2
DDM	dodecyl- β-D-maltopyranoside
ep	error prone
epPCR	error prone PCR
FACS	fluorescence-activated cell sorter
FI	fluorescence intensity
GFP	green fluorescent protein
GPCR	G-protein coupled receptor
GRAFS	glutamate rhodopsin adhesin frizzled/taste2 secretin
H_1	histamin receptor type 1
HMFM	Hogness Modified Freezing Medium
IMAC	immobilized metal-ion affinity chromatography
IPTG	isopropyl- β-D-thiogalactopyranoside
K_d	equilibrium dissociation constant
LBA	ligand binding assay
MBP	maltose binding protein
MFI	mean fluorescence intensity
MSA	multiple sequence alignment
NCBI	National Center for Bioinformatic Information
NT	neurotensin
NTR1	neurotensin receptor type 1
OD	optical density
OM	outer membrane
PCR	polymerase chain reaction
PDB	Protein Data Bank

RMSD	root mean square deviation
SASA	solvent accessible surface area
SCAM	substituted-cysteine accessibility method
SI	stability index
SP	substance P
SP-OG	Substance P Oregon Green
TACR1	tachykinin receptor type 1
TEV	tobacco etch virus
TrxA	thioredoxin A
wt	wild-type

Curriculum vitae

Name	Dodevski
Fist Name	Igor
Date of Birth	April 25th 1978
Place of origin	Laufen BL (Heimatort)

Education

1990 – 1997 **Regionales Gymnasium Laufental-Thierstein, Switzerland**
Eidgenössische Matura Typus B

1998 – 2000 **University of Basel, Switzerland**
Grundstudium in Molekularbiologie

2000 – 2003 **Ecole Supérieure de Biotechnologie de Strasbourg, France**
Master of Science (Diplombiotechnologe; Diplôme d'Ingénieur en Biotechnologie) and Diplôme d'Etudes Approfondies (DEA) en biologie cellulaire et moléculaire

2003 **University of California San Francisco, USA**
Diploma Thesis in the group of Prof. Patricia C. Babbitt, Ph.D.
Title: "*Bioinformatic analysis of the enolase superfamily to develop and apply a protocol for identifying misannotated sequences in public databases"*

2004 – 2010 **University of Zürich, Dept. of Biochemistry, Switzerland**
Graduate studies at the Life Science Zürich Graduate School
Dissertation in the group of Prof. Dr. Andreas Plückthun
Title: *"Directed evolution of G-protein coupled receptors for expression and stability"*